

Global Ensemble Predictions of 2009's Tropical Cyclones

Initialized with an Ensemble Kalman Filter

Thomas M. Hamill¹, Jeffrey S. Whitaker¹,
Michael Fiorino,² and Stanley G. Benjamin²

¹ *NOAA Earth System Laboratory, Physical Sciences Division, Boulder, Colorado*

² *NOAA Earth System Laboratory, Global Systems Division, Boulder, Colorado*

Submitted to *Monthly Weather Review*

as part of the 3rd THORPEX International Science Symposium Special Collection

8 April 2010

Corresponding author address:

Dr. Thomas M. Hamill
NOAA/ ESRL / PSD
R/PSD1, 325 Broadway
Boulder, CO 80305-3328
Phone: (303) 497-3060
Fax: (303) 497-6449
E-mail: tom.hamill@noaa.gov

ABSTRACT

Verification was performed on ensemble forecasts of 2009 Northern Hemisphere summer tropical cyclones (TCs) from two experimental global numerical weather prediction ensemble prediction systems (EPSs). The first model was a high-resolution version (T382L64, ~45 km grid spacing at 30°N, 64-level) of the National Centers for Environmental Prediction's (NCEP) Global Forecast System (GFS). The second model was a 30-km version of the experimental NOAA Earth System Research Lab's Flow-following finite-volume Icosahedral Model, or "FIM." Both models were initialized with the first 20 members of a 60-member ensemble Kalman filter (EnKF) using the T382L64 GFS. The GFS/EnKF assimilated the full observational data stream that was normally assimilated into the NCEP operational Global Statistical Interpolation (GSI) data assimilation, plus human-synthesized "observations" of tropical cyclone central pressure and position produced at the National Hurricane Center and the Joint Typhoon Warning Center. The forecasts from the two experimental ensembles were compared against four operational EPSs: (1) the operational T399L62 European Centre for Medium-Range Weather Forecasts (ECMWF) global ensemble; (2) the T126L28 operational NCEP ensemble; (3) the 0.9-degree, 28-level operational Canadian Meteorological Centre (CMC) ensemble; and (4) and the 1.25x0.83 degree gridded United Kingdom Met Office (UKMO) ensemble. Forecast tracks and intensities were validated relative to post-season "best-track" estimates produced by the National Hurricane Center and the Joint Typhoon Warning Center.

The errors of GFS/EnKF ensemble track forecasts were competitive with those from the ECMWF ensemble system, and the overall spread of the ensemble tracks was consistent in magnitude with the track error. Both experimental EPSs had much lower errors than the operational NCEP, UKMO, and CMC EPSs, but the FIM/EnKF tracks were somewhat less accurate than the GFS/EnKF. The ensemble forecasts often were stretched in particular directions, and not necessarily along- or across track. The better-performing EPSs provided useful information on potential track error anisotropy. While the GFS/EnKF initialized relatively deep vortices by assimilating the TC central pressure estimate, the model storms filled during the 24 h. Other forecast models also systematically under-estimated TC intensity (max forecast surface wind speed). The higher-resolution models generally had less bias.

Analyses were conducted to try to understand whether the additional central pressure observation, the EnKF, or the extra resolution was most responsible for the decrease in track error of the experimental GEFS/EnKF over the operational NCEP. The assimilation of the additional TC observations produced only a small change in deterministic track forecasts initialized with the GSI. The T382L64 GFS/EnKF ensemble was used to initialize a T126L28 ensemble forecast to facilitate a comparison with the operational NCEP system. The T126L28 GFS/EnKF EPS track forecasts were dramatically better than the NCEP operational, suggesting the positive impact of the EnKF, perhaps through improved steering flow.

1. Introduction

The accuracy of official National Hurricane Center tropical cyclone (TC) track forecasts has improved over the past several decades (Rappaport et al. 2009). In part due this can be attributed to the general improvements in numerical weather prediction (NWP) models, such as increased resolution, improved methods of initialization, more realistic physical parameterizations, and the availability of a greater number of skillful models and ensemble systems for generating consensus forecasts. Meanwhile, improving the accuracy of intensity forecasts has proven much more difficult (Bender and Ginis 2000; Krishnamurti et al. 2005, Rogers et al. 2006, Li and Pu 2008, Rappaport et al. 2009).

NOAA's Hurricane Forecast Improvement Project (HFIP; <http://tinyurl.com/ykdk6qv>) was initiated to improve the accuracy and reliability of hurricane forecasts, to extend the forecast lead time for hurricane forecasts with increased certainty, and to increase confidence in hurricane forecasts. HFIP has a specific objective to reduce track and intensity errors by 50 percent by 2019 relative to the 2009 baseline.

An important component of HFIP are the development of improved methods for global ensemble prediction systems (EPSs)¹. Suppose that all sources of forecast error such as chaotic growth of initial-condition errors and the cumulative effects of model uncertainty and sampling error could be properly addressed in an ensemble prediction system. Then, the actual hurricane and the forecast ensemble members

¹ Ensemble predictions from regional models will not be considered in this article.

could be able to be considered random draws from the same underlying probability distribution (Hamill 2001). As ensemble systems are improved and begin to approach this ideal, the utility of hurricane forecast products could also be improved dramatically. For example, the mean or mode of the track of the ensemble distribution will then provide on average a more accurate forecast than any individual member, especially at longer leads, thus contributing to the HFIP goal of a reduction in track error. The diversity of track forecasts and maximum surface wind-speed forecasts (TC intensity) could be used to quantitatively assess risk and make more appropriate and earlier decisions about coastal evacuations. The current state of the art of TC forecasts from global operational ensemble systems was recently described in Majumdar and Finnochio (2010).

Two technological improvements that may contribute to improved global ensemble predictions are the use of higher-resolution global models and the use of advanced methods for data assimilation and ensemble initialization. Regarding the former, many experiments in global and regional forecast models have shown that the realism of tropical cyclones in models is improved as the horizontal and vertical resolutions are refined (e.g., Krishnamurti et al. 1989; Lam 2001, Zhang and Wang 2003; Bender et al. 2007, Davis et al. 2008, Fiorino 2009; Gentry and Lackmann 2010). Since the mean radius of maximum winds in tropical cyclones is ~ 55 km (Kimball and Mulekar 2004) and that ~ 4 grid points (Grasso 2000) to 7-10 grid points (Skamarock 2004, Fiorino and Elsberry 1989ab) per wavelength are needed to resolve a feature minimally, this suggests that a model with a 137-km grid spacing (at 30° N) such as the 2009 version of the National Center for

Environmental Prediction (NCEP) Global Ensemble Forecast System (GEFS) will not be able resolve the important tropical cyclone dynamics². Even a global model with 43-km grid spacing cannot be expected to produce realistic tropical cyclone vortices, and this grid spacing was the approximate state of the art in 2009 global ensemble prediction systems.³ Still, some improvement in tropical cyclone representation in the models as well as the accuracy of track forecasts may be possible as resolution is increased to from ~ 137 to ~ 43 km; tropical cyclones may appear more realistic, especially the larger ones, and the interaction of the vortex with the steering flow may be handled more accurately.

Concerning advanced data assimilation methods, four-dimensional variational data assimilation (“4D-Var;” Le Dimet and Talagrand 1986; Courtier et al. 1994; Rabier et al. 2000, Rawlins et al. 2007) has been a standard method for data assimilation at many operational centers worldwide. However, the typical 4D-Var does not directly produce an ensemble of initial conditions that sample the desired analysis uncertainty, so some other method is needed for generating an ensemble of initial conditions. Operational methods in 2009 include adding perturbations to the control via singular-vector techniques (Barkmeijer et al. 1998, 1999), the ensemble transform technique (Wei et al. 2008), the ensemble transform Kalman filter

² This resolution model may still be able to effectively steer whatever TC-like disturbance is there, so such models are far from being useless.

³ The highest-resolution and most accurate ensemble system (in most respects) was from the European Centre for Medium-Range Weather Forecasts (ECMWF), which employed an ensemble prediction system with horizontal T399 resolution, or approximately 43 km at 30°N. In early 2010, the resolution was increased to T639, or ~ 27 km at 30°N.

(Bishop et al. 2001, Wang and Bishop 2003) and the local ensemble transform Kalman filter (Hunt et al. 2007, Bowler et al. 2009).

Recently, another advanced data assimilation and ensemble initialization technique has emerged: the ensemble Kalman filter or “EnKF” (Evensen 1994, Houtekamer et al. 1998, Lorenc (2003), Hamill (2006), Ehrendorfer (2007), and Evensen (2009)). Ensembles of short-term forecasts are used to estimate the background-error covariances in the update step, where forecasts are adjusted to newly available observations. This update step produces an ensemble of analyses that are specifically constructed to sample the analysis uncertainty, and these analyses are used as initial conditions in subsequent short-range forecasts for the next update step. Longer-lead ensemble predictions may be generated directly from these initial conditions as well.

The NWP community has recently tested the EnKF and its variants with real data, with many studies showing positive impact (Whitaker et al. 2004, 2008; Houtekamer et al. 2005, 2009; Houtekamer and Mitchell 2005; Compo et al. 2006; Miyoshi and Yamane 2007; Meng et al. 2008ab; Torn and Hakim 2008, 2009; Wang et al. 2008; Szunyogh et al. 2008; Zhang et al. 2009, Aksoy et al. 2009; Buehner et al. 2009ab). Operationally, the Canadian Meteorological Centre now uses the EnKF to produce the perturbed initial conditions around its 4D-Var analysis.

Because the EnKF produces flow-dependent estimates of the background-error covariances, it may be particularly helpful with TC initialization, where the isotropic background-error covariances that are commonly used in variational

methods are particularly inappropriate (Fig. 1). Recent studies of the EnKF applied to tropical cyclones include Chen and Snyder (2007), Zhang et al. (2009), and Torn and Hakim (2009). The EnKF may have an additional benefit in that the background-error covariance model in variational systems may implicitly reflect balances between the mass and wind fields that are tuned for the mid-latitudes and less appropriate in the tropics. Mass-wind balances in the EnKF covariances are set directly by the ensemble-estimated covariances, which may have large-scale mass-wind balance in the extratropics but different relationships in the tropics.

To date, however, no experiments have been performed with an EnKF applied in a high-resolution global forecast model with a full observational data set. The combined application of a high-resolution global model and the EnKF may be particularly beneficial to TC ensemble forecasting. As model resolution is increased and the fidelity of TCs in the model is improved, the error covariances around the hurricane may be more appropriately estimated in the EnKF; this may produce reduced-error state estimates and better forecasts for the next assimilation cycle.

Given the conceptual appeal of a high-resolution EnKF, an experiment was performed in the summer of 2009 to run the EnKF with global models at resolutions comparable to ECMWF's benchmark, in our case with a T382L64 model (~45 km grid spacing at 30°N) and a 30-km implementation of the FIM model. This paper describes the results of these experiments, and specifically the characteristics of ensemble predictions from this system and how they compare with operational ensemble predictions from other centers. The operating hypothesis is that the

combination of the EnKF and a higher-resolution global model will act synergistically to produce a significant improvement in TC forecast guidance.

Accordingly, the rest of the manuscript is organized as follows. Section 2 will provide a description of the forecast models tested, the data assimilation scheme used, the observations that were assimilated, and the methods for data analysis. Section 3 provides results, and section 4 provides concluding remarks.

2. Methods and data.

a. Forecast models

The experiments were designed to focus primarily on the performance of two experimental EPSs, both initialized by the GFS-based EnKF. The first was the NCEP Global Forecast System Model (GFS), a global spectral model with a hybrid sigma-pressure vertical coordinate. In this experiment the GFS was used at T382L64 resolution (wavenumber 382 in triangular truncation, or approximately 45 km at 30° latitude, and 64 vertical levels). The model employs the primitive equations with the following primary prognostic variables: vorticity, divergence, logarithm of surface pressure, specific humidity, virtual temperature, and total cloud condensate. The treatment of vertical mixing, including the planetary boundary layer, was based on Hong and Pan (1996) and Troen and Mahrt (1986). Short-wave radiation followed Chou (1992), Chou and Lee (1996), Chou et al. (1998), while long-wave radiation implements the Rapid Radiative Transfer Model (RRTM) of Mlawer et al. (1997). Penetrative convection was developed by Pan and Wu (1994),

based on a Grell (1993) implementation of Arakawa and Schubert (1974). The effect of non-precipitating shallow clouds is incorporated following Tiedtke (1983). Cloud formation on resolved scales was treated according to Zhao and Carr's (1997) modification of Sundqvist et al. (1989). Attenuation of gravity waves propagating into the mesosphere was accomplished with Rayleigh drag. A description of the GFS model is available from the NCEP Environmental Modeling Center (EMC), with changes as of 2003 described at www.emc.ncep.noaa.gov/gmb/moorthi/gam.html.

The second experimental EPS used the NOAA Earth System Research Laboratory's "FIM" model, the hydrostatic, Flow-following, finite-volume Icosahedral Model. The main difference between the FIM and the GFS was the dynamical core; otherwise, the physical parameterizations were adapted from the GFS. The horizontal grid for the dynamical core was a spherical geodesic grid, and the Arakawa-A implementation was used (no staggering of temperature and wind components; Arakawa and Lamb 1977). The vertical coordinate was hybrid sigma-isentropic, and finite-volume horizontal transport was used. The primary prognostic variables were pressure, potential temperature, u - and v - wind components, water vapor mixing ratio, and using GFS physics cloud condensate. Further information on the FIM model can be found in Lee and MacDonald (2009), Bleck et al. (2010), and at <http://fim.noaa.gov>. The FIM ensemble forecasts in this study were conducted at G8 resolution, or ~ 30 km grid spacing uniform globally, and 64 vertical levels were used, the same as the GFS. The FIM ensemble forecasts were initialized with the GFS-EnKF, just as were the GFS T382 ensemble forecasts.

Comparisons are also provided against the track forecasts from other operational global EPSs. Specifically, the experimental “GFS/EnKF” track forecasts were compared against the 2009 operational EPS forecasts from the Canadian Meteorological Centre (CMC), the European Centre for Medium-Range Weather Forecasts (ECMWF), the UK Met Office (UKMO), and the NCEP Global Ensemble Forecast System (NCEP).

The CMC EPS used the Global Environmental Multiscale Model (GEM), a hydrostatic primitive equation model with a terrain-following pressure vertical coordinate. Further documentation on the GEM model can be found at <http://tinyurl.com/yl3cftb> and in Charron et al. (2010). The CMC ensemble system used a horizontal computational grid of 400x200 grid points, or approximately 0.9 degrees, and 28 vertical levels. The ensemble perturbations around its 4D-Var control analysis were initialized with the perturbations of ensemble Kalman filter analyses about its mean EnKF analysis, following Charron et al. (2009) and Houtekamer et al. (2009) and references therein. The 20 forecast ensemble members used a variety of perturbed physics; changing gravity wave drag parameters, land-surface process type, condensation scheme type, convection scheme type, shallow convection scheme type, mixing-length formulation, and turbulent vertical diffusion parameter. More details on these are provided at <http://tinyurl.com/ygwdt2v>.

The ECMWF EPS used the ECMWF Integrated Forecast System (IFS) model, versions 35r2 (prior to 8 Sep 2009) and 35r3 (after 8 Sep 2009). Model resolution

was T399L62 for both versions, or about 43-km grid spacing at 30° latitude; details on the IFS are provided at www.ecmwf.int/research/ifsdocs/. The changes to the ensemble stochastic treatments in the 8 Sep 2009 implementation are described in Palmer et al. (2009). The ensemble was initialized with a combination of initial-time and evolved total-energy singular vectors (Buizza and Palmer 1995, Molteni et al. 1996, Barkmeijer et al. 1998, 1999, Leutbecher 2005) and utilized stochastic perturbations to physical tendencies. Recent performance of the ECMWF IFS for TC prediction was described in Fiorino (2009), and an overview of the ensemble system was provided in Buizza et al. (2007) and references therein. For consistency with the analysis of other EPSs, only the first 20 perturbed members were used here.

The UK Met Office (UKMO) ensemble system was “MOGREPS,” the Met Office Global and Regional Ensemble Prediction System. TC track forecasts from this system came from its global component, which was described in Bowler et al. (2008, 2009). The global system was run at a resolution of 1.25° longitude and 0.83° latitude on a regular latitude-longitude grid. 38 vertical levels were employed. Initial condition perturbations were generated from an implementation of the ensemble transform Kalman filter (Hunt et al. 2007, Bowler et al. 2009). The mean initial state was generated from the UKMO 4D-Var system (Rawlins et al. 2007). The model included a parameterization of one type of model uncertainty via its stochastic kinetic-energy backscatter scheme, following Shutts (2005).

The NCEP global ensemble system in 2009 utilized the GFS forecast model at T126L28, a grid spacing of ~ 137 km at 30° latitude. The control initial condition was produced by the T382 GSI analysis, which included a procedure for the relocation of vortices (Liu et al. 2000). The analysis was then degraded to T126L28 resolution. Perturbed initial conditions were generated with the ensemble transform technique of Wei et al. (2008). The 2009 implementation did not include any parameterizations of model uncertainty.

b. Data assimilation system and observations used.

The primary data assimilation system and method for initialization of the experimental hurricane ensemble forecasts was a 60-member ensemble square-root filter (Whitaker and Hamill 2002). This is a deterministic version of the ensemble Kalman filter (hereafter, “EnKF”) and used the T382L64 GFS as the forward model. In several instances below, EnKF analyses were compared to those provided by an equivalent resolution version of the operational NCEP Grid-point Statistical Interpolation (GSI) 3-D variational system (Wu et al. 2002, Kleist et al. 2009). A recent implementation of the EnKF in the GFS was described more fully in Whitaker et al. (2008). Several changes were made to this EnKF algorithm for the 2009 experiment. The NCEP GSI system was applied to the prior forecasts for each ensemble member (saved hourly in the 3-6 hour forecast window) to compute the forward operator (the full variational minimization was skipped). An online adaptive radiance bias correction algorithm developed by T. Miyoshi (personal communication) was included in the EnKF using the same predictors as used in the

GSI. Additionally, the fast parallel algorithm of Anderson and Collins (2007) was used to calculate the EnKF increment. Updates to new observations were performed every 6 h and included data in a 3- to 9-hour forecast window. Covariance localization was applied in the horizontal that forced increments to zero 1500 km from the observation location. Localization was also applied in the vertical, forcing increments to zero at 1.5 scale heights (computed as the natural log of the ratio of the pressure of the model level divided by the pressure assigned to the observation location). The reference pressure for radiance observations was set to the maximum in the weighting function for each instrument and channel. System error (including model error) was parameterized with a combination of multiplicative and additive inflation.

Since our version of the EnKF used the GSI's quality control algorithms and forward observational operators, the full observational data stream that was ingested into the operational GSI was also ingested into the EnKF, including surface observations, rawinsondes, cloud-track winds, aircraft observations, satellite-based Global Positioning System (GPS) radio occultations, and satellite radiances. Satellite radiances were assimilated from the High-resolution infrared sounder (HIRS), the Advanced Microwave Sounding Unit (AMSU) A and B channels, Atmospheric Infrared Sounder (AIRS), Microwave Humidity Sounder (MHS), Geostationary Orbiting Environmental Satellite (GOES) radiances, ozone retrievals, and Special Sensor Microwave/Imager (SSM/I). Additionally, the operational 'working best track' estimates of TC mean-sea level pressure and position, called "TCVitals," were assimilated (Guard et al. 1992, Rappaport et al. 2009).

c. Configuration of the experimental ensemble forecasts.

The experimental ensemble Kalman filter system was cycled between 0000 UTC 1 Jul 2009 to 0000 UTC 28 Sep 2009. However, allowing for a spin-up period and because of intermittent outages, experimental ensemble forecasts were computed only with initial conditions between 20 Jul 2009 and 28 Sep 2009, and only from 0000 UTC initial conditions. Moreover, both experimental and operational ensembles were not uniformly available during this period (Table 1). Whereas 60 forecasts were cycled for the data assimilation, only 21 ensemble forecasts were conducted, consisting of the control analysis and the first 20 perturbed members. For all experiments below, we will consider the statistics from the 20 perturbed members, neglecting the control forecast so that the ensemble statistics can be computed with members that are approximately exchangeable in their statistical properties. Forecasts statistics were computed only to 5 days lead-time.

d. Methods of data analysis.

A set of rules and metrics was established to facilitate an equitable comparison between the ensemble prediction systems. Unlike operational facilities, our experimental forecasts were not uniformly available on all dates during the season, and for most dates the experimental ensemble predictions were generated only from 0000 UTC analyses. Hence, to ensure a fair head-to-head or “homogeneous” comparison (at the expense of limiting the sample size) we established the following criteria for including a particular tropical cyclone in the verification sample: (a) in comparisons between two models, both forecasts must

have been available; (b) the cyclone must have been of at least tropical depression strength at the initial time of the forecast, as specified in the TCVitals “best track” reports. More specifically, the storm must have been classified as a tropical cyclone, i.e., a tropical depression (TD), tropical storm (TS), hurricane (HU), typhoon (TY), super typhoon (ST), tropical cyclone (TC), or inland (IN) at the initial time; (c) the observed TC must have continued to be of at least tropical depression strength at the particular lead time being evaluated; (d) the cyclone must have been tracked in the ensemble forecasts by at least 40 percent of the members for each of the systems; (e) even if the EPS generated more than 20 ensemble-member forecasts, only the first 20 perturbed member forecasts were utilized, and the control member was not considered. Several of the figures below will present “non-homogeneous” data, where restriction (a) above was not enforced.

The most commonly used metrics were the average absolute track error $\bar{E}(t)$ for time t , i.e., the great circle distance between forecast and verifying position), and the average spread $\bar{S}(t)$ of the track forecasts. Suppose we have m samples, and the absolute error $E_i(t)$ represents the absolute error in km of the ensemble-mean track forecast for the i th of m samples. The average absolute track error is thus defined as:

$$\bar{E}(t) = \frac{\sum_{i=1}^m E_i(t)}{m} . \quad (1)$$

A calibrated ensemble system should also provide consistency between the error and spread, presuming the two are measured in similar ways. Accordingly, a consistent way of measuring spread $S_i(t)$ for the i th sample, is

$$S_i(t) = \frac{\sum_{j=1}^n D_{i,j}}{n} , \quad (2)$$

where $D_{i,j}$ is the distance of the j th of n members from the ensemble-mean position.

The average spread is defined as

$$\bar{S}(t) = \frac{\sum_{i=1}^m S_i(t)}{m} . \quad (3)$$

Note that model performance will only be evaluated with statistics across all basins; samples from the western, central, and eastern Pacific, the Atlantic, and the Indian Oceans will all be considered together. Because of the limited sample size, basin-by-basin statistics would be untrustworthy.

3. Results.

a. Review of the 2009 Northern Hemisphere hurricane season.

The 2009 hurricane season was relatively quiet in the Atlantic Basin during the period of the experiment, with no land-falling hurricanes (Fig. 2). The most significant hurricane of the season was Bill, which reached category 4 on the Saffir-Simpson scale (<http://www.nhc.noaa.gov/sshws.shtml>) on 19 Aug 2009 but did not affect the US coastline. There were several significant storms in the eastern (Fig. 3) and western Pacific (Fig. 4). In the eastern Pacific, the compact hurricane Felicia

strengthened to a category 4 storm briefly on 5 Aug 2009, which made it the strongest storm in the basin since 2006. Hurricane Jimena made landfall as a weak category 2 hurricane near Cabo San Lazaro, Mexico on 2 Sep 2009. Long-lived Guillermo was briefly a category 3 hurricane on 15 Aug 2009. Meanwhile, there were several very significant typhoons in the western Pacific. In terms of human impact during the period of the experiment, the most damaging was Typhoon Morakot, which made landfall on the northeast coast of Taiwan as an approximately Category 1 storm on 7 Aug 2009. However, its immense circulation created a sustained period of upslope rainfall in the mountains of southern Taiwan, with a peak rainfall of 1397 mm (~55 in; http://disc.sci.gsfc.nasa.gov/gesNews/typhoon_morakot). This extreme precipitation triggered massive mudslides, including one slide that buried the entire village of Xiaolin, killing approximately 500 people. Damages were estimated at over 3 billion US dollars. The storm also caused extensive flooding and damage in China and the Philippines. Tropical storm Ketsana (called “Ondoy” in the Philippines) also produced extreme rainfall. Metro Manila received 341 mm between 8 am and 2 pm local time on 26 Sep 2009 (<http://tinyurl.com/ycv3pvp>). There were also a very few number of cyclones in the Indian Ocean as well (not shown).

b. Track comparisons between models.

Figure 5 shows mean error and spread statistics from the experimental GFS/EnKF track forecasts relative to the experimental FIM/EnKF forecasts, ECMWF,

UKMO, CMC, and NCEP. The GFS/EnKF forecasts greatly outperformed the NCEP operational system, with its GSI/ET initialization and roughly 3 times coarser horizontal resolution. The spreads of NCEP operational forecasts were much smaller than their errors, while spreads and errors were similar in magnitude in the GEFS/EnKF. The GFS/EnKF forecasts were competitive with the ECMWF forecasts and smaller in error at the longest lead, though the difference was barely significant at the 5% level. The GFS/EnKF was clearly better than the CMC ensemble, though the CMC system did have a general consistency of spreads and errors. The UKMO ensemble had higher errors and much less spread than error. Finally, the experimental FIM/EnKF ensemble was not quite as accurate as the GFS/EnKF, though it outperformed all operational ensembles except the ECMWF and had a general consistency between spreads and errors. Taken together, these results suggest that the increased resolution, the extra TCVitals observations, and/or the advanced assimilation method could dramatically improve operational hurricane track prediction (see section 3d below for the relative benefits of observations, resolution, and the data assimilation).

Consider briefly the track forecasts from some of the more important storms of 2009. Figure 6 shows ensemble forecasts for Hurricane Bill initialized at 0000 UTC 19 Aug 2009. The most notable feature in the various track forecasts is that the ensemble forecasts were slightly too slow in forecasting the recurvature of Bill to varying extents, and the CMC and UKMO forecasts exhibited a westward forecast bias. Figure 7 shows the tracks of Typhoon Morakot initialized at 0000 UTC 5 August 2009. The experimental GEFS/EnKF cyclone forecast tracks were quite

accurate prior to landfall on Taiwan, though they exhibit a slight northward bias. The NCEP tracks were too slow, and most of the tracked forecast cyclones never made landfall on Taiwan. ECMWF were also quite skillful leading up to landfall. CMC forecasts were also reasonable prior to landfall, though CMC forecasts had a slight northward bias and very large spread subsequent to landfall. UKMO forecasts were too far north and too fast. Finally, Fig. 8 shows hurricane Jimena forecasts initialized at 0000 UTC 30 August 2009. The forecasts generally had a large westward bias, keeping the predicted cyclones unrealistically off shore. However, the observed track was generally within the span of the ensemble in the GEFS, ECMWF, and CMC ensemble systems.

Was there any positive relationship between the spread of the positions and the absolute error of the positions? Whitaker and Lough (1998) showed that potential spread-error correlations should be larger when there are greater day-to-day and case-to-case variations in the spread. Accordingly, Fig. 9 shows spread-error relationships for each of the ensemble systems for day+3 forecasts. There were very modest correlations between spread and error, from 0.12 for the GFS/EnKF to 0.38 for the NCEP system. Was this pointing out a deficiency of the ensemble prediction systems, or was the potential for a spread-error correlation limited due to the lack of variations in spread? Suppose that the ensemble system were “perfect” in that true state was a random draw from one of the tracked forecast ensemble members (Buizza 1997). How much correlation can be expected then? Figure 10 presents spread-error data under such a construct. As shown, correlations of ~ 0.32 to 0.58 could be expected under perfect-model assumptions;

the general lack of correlation was thus likely to be attributable in part to deficiencies in the ensemble system designs rather than the modest variations in forecast spread.

Figures 6-8 showed that the spread of the ensemble was often concentrated in one direction more than another, and not necessarily along- or across track. We now consider whether the elliptical nature of the forecast errors was realistic. We examined this by considering the average projections of the forecast errors onto the eigenvectors of the forecast-error covariance matrix (Fig. 11) and how this compares to the average projection of individual forecast members onto the eigenvectors. Let us assume that we have a vector of length nt (the number of ensemble members that tracked the storm) denoting the scaled ensemble deviations about the ensemble mean. For longitude λ , this vector was

$$\mathbf{x}'_{\lambda} = (x_{\lambda(1)} - \bar{x}_{\lambda}, \dots, x_{\lambda(nt)} - \bar{x}_{\lambda}) / (nt - 1)^{1/2}, \quad (4)$$

where \bar{x}_{λ} denotes the ensemble-mean longitude and $x_{\lambda(i)}$ the longitude of the i th member forecast. Similarly, for latitude ϕ we had the vector

$$\mathbf{x}'_{\phi} = (x_{\phi(1)} - \bar{x}_{\phi}, \dots, x_{\phi(nt)} - \bar{x}_{\phi}) / (nt - 1)^{1/2}. \quad (5)$$

Let

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}'_{\lambda} \\ \mathbf{x}'_{\phi} \end{bmatrix}. \quad (6)$$

We then formed the sample forecast-error covariance matrix \mathbf{F} and obtained its eigenvalue decomposition as:

$$\mathbf{F} = \mathbf{X}\mathbf{X}^T = \mathbf{S}\mathbf{\Lambda}\mathbf{S}^{-1} = \mathbf{S}\mathbf{\Lambda}\mathbf{S}^T = (\mathbf{S}\mathbf{\Lambda}^{1/2})(\mathbf{S}\mathbf{\Lambda}^{1/2})^T, \quad (7)$$

where \mathbf{S} denote the eigenvectors of the covariance matrix and $\mathbf{\Lambda}$ the associated eigenvalues. Eigenvalues and eigenvectors were ordered so that the leading eigenvector was associated with the larger eigenvalue. $\mathbf{S}^{-1} = \mathbf{S}^T$ since the covariance matrix was real and symmetric (Strang 1988, p. 296). The square roots of the eigenvalues provide one measure of spread (the square-root of variance) along the direction of each of the eigenvectors.

For some given case, let \mathbf{E} denote the ensemble-mean forecast error vector, the difference between the ensemble mean and the observed positions (see Fig. 11 for a schematic). Similarly, let the vector \mathbf{X}_i denote the i th forecast member's vector deviation from the mean. Let $\langle \cdot \rangle$ denote the average over many cases, and let $\langle \langle \cdot \rangle \rangle$ denote the average over many cases and all available ensemble members. Ideally, the average absolute projection of the ensemble-mean forecast error vector \mathbf{E} onto the eigenvectors, $\langle |\mathbf{E} \cdot \mathbf{S}_1| \rangle$ and $\langle |\mathbf{E} \cdot \mathbf{S}_2| \rangle$, should be consistent with the average absolute projections of any ensemble member's vector deviation from its mean onto the eigenvectors, $\langle \langle |\mathbf{X}_i \cdot \mathbf{S}_1| \rangle \rangle$ and $\langle \langle |\mathbf{X}_i \cdot \mathbf{S}_2| \rangle \rangle$, respectively. Figure 12 indicates the extent to which this occurred for each forecast ensemble. Consider panel (a), the GFS/EnKF forecasts. The solid blue line indicates the average ensemble projection onto the first eigenvector $\langle \langle |\mathbf{X}_i \cdot \mathbf{S}_1| \rangle \rangle$ as forecast lead increased. Consistent with the general increase with spread with time, the magnitude of this

increased from ~ 60 km at day 0.5 to 220 km at day 5. Meanwhile, the average projection onto the second eigenvector, $\langle \langle \mathbf{X}_i \cdot \mathbf{S}_2 | \rangle \rangle$, the solid red curve in Fig. 12(a), increased from ~ 40 km to ~ 190 km. Now consider the projection of the error onto the associated eigenvectors. For the first eigenvector (the dashed blue curve), the projection of the error was quite consistent with the projection of the forecast, though at the longest leads the projection of error was smaller, indicating a deficiency of spread. For the second eigenvector, between days 2.5 and 4.5, the average projection of error was slightly larger than the average projection of the ensemble members, indicating a modest deficiency of spread in this direction. Together, these suggest that a more ideal ellipse would have been more isotropic. However, considering the two dashed curves, there was somewhat more error in the direction of the leading eigenvector than in the direction of the trailing, indicating some ability for the ensemble to distinguish the directions in phase space where errors are more pronounced.

Let's consider the characteristics of some of the other ensemble systems. For the NCEP system, in both eigendirections there was a greater projection of error onto the eigenvectors, indicating a deficiency of spread. ECMWF had the best consistency between error and spread, at least after day 2. The CMC forecasts were only slightly under-spread, with better consistency along the leading eigenvector than along the trailing one. The UK Met Office ensemble, like NCEP, was under-spread, and the FIM ensemble appeared to have good consistency through about day 3.5.

c. Intensity forecasting.

Figure 13 shows that the global forecast models with grid spacings of > 43 km and above did not provide realistic maximum wind-speed forecasts, though the finer-resolution models generally had less bias. While the GFS/EnKF and the FIM/EnKF model started out with relatively small analysis biases of ensemble-mean wind speed, during the first 24 h they rapidly increased their bias. The ECMWF model started with a larger analysis bias (ECMWF did not assimilate the TCVitals observations), but unlike the GFS/EnKF, their biases did not grow during the first day. The lower-resolution models, NCEP and CMC, had larger biases (wind speed information was not available from the UKMO system).

What caused the rapid bias change of the GFS/EnKF during the first day? Was this due to the model, the data assimilation method, and/or the observations used? To examine this, we considered day+1 deterministic forecasts generated from the ensemble-mean initial condition and compared them against deterministic forecasts generated from a parallel run of the T382L64 GFS. However, this time the data assimilation was performed using the GSI with the full operational data stream, plus TCVitals (Fig. 14). This will be called “GSI-Parallel.”

Figure 14a shows that the GFS/EnKF change in wind speed bias in the first 24 h strongly depended on the initial intensity of the cyclone; wind speeds in stronger cyclones generally decreased more. Interestingly, the GSI-Parallel did not show much if any intensity change on average between day 0 and day 1 (Fig. 14b).

This can be understood by considering the initial analyzed intensities (Figs. 14cd). Whereas the GFS/EnKF produces analyses that were similar in intensity to the observed, the GSI-Parallel analyzed vortices that were much less intense than observed. Figure 15 shows that in an example from hurricane Ike in 2008, the initialized vortex in the EnKF was much deeper and narrower than the vortex from the GSI. We suspect that the T382 GFS was simply too coarse a resolution to support such a narrow, deep vortex and consequently filled rapidly during the first few hours of the forecast.

d. The relative benefits of observations, resolution, and the data assimilation.

The experimental GFS/EnKF provided a large improvement in the skill of track forecasts relative to the 2009 NCEP operational system (Fig. 5a). There were at least three possible reasons for the improvement: the threefold increase in model resolution, the additional TCVitals observations, and the method for initializing the ensemble. To attempt to quantify the effect of the data assimilation and the TCVitals observations, Fig. 16a shows the track errors for deterministic forecasts from the T382 GFS/EnKF and two versions of the T382 GFS initialized with the GSI. The first version was the “GSI parallel,” which included the assimilation of TCVitals⁴, and the second was the “GSI operational,” which did not. There was a slight decrease in the error of the track forecasts from the GSI parallel relative to GSI operational, though the difference was not statistically significant. We conclude that the TCVitals observations did not markedly affect the performance of GSI-based

⁴ This version became operational in March 2010.

forecasts, though it is possible that the EnKF, which drew more to the observations (Figs. 14 and 15), would have been more affected. The GFS/EnKF did have substantially lower track errors, suggesting that the EnKF assimilation method produced an improvement in the forecast. The ensemble averaging effect apparently did not have a large impact on the skill of the tracks (Fig. 16b). Interestingly, the track errors were only modestly larger when the GFS/EnKF forecasts were conducted at T126L28 resolution, upscaled from T382L64 (Fig. 17a). However, these T126L28 ensemble track forecasts initialized with the EnKF were much lower in error than the operational T126L28 ensemble forecasts (Fig. 17b). This result suggests that the additional resolution may be critical to improving the analysis from the EnKF, but thereafter the increased forecast resolution was less critical. Perhaps the increased accuracy of EnKF forecasts resulted from an improvement in the analyses and forecasts of tropical steering winds for the tropical cyclones (Fig. 18).

5. Conclusions.

There is an urgent need for improved tropical cyclone forecasts, both for track and intensity. There are several possible routes to improve these forecasts, including assimilating new types of observations, improving the data assimilation, and increasing the resolution of the forecast model substantially. In this manuscript we attempt to determine if these three would substantially improve the ensemble predictions of tropical cyclones for the 2009 Northern Hemisphere summer. The new source of observational data was the human-synthesized

“TCVitals” observations of cyclone minimum sea-level pressure and position. The new data assimilation and ensemble initialization methodology was an EnKF. The resolution increase was approximately threefold, from 2009 NCEP operational T126L28 (~137 km at 30° N) to an experimental T382L64 (~ 45 km at 30° N). There was also reason to believe these three improvements might work synergistically. For example, the potentially more appropriate background-error covariances from the EnKF may permit a more appropriate use of TCVitals observations than the current operational 3D-Var scheme at NCEP. Also, the higher resolution might result in an ensemble that generated more appropriate background-error covariances in the EnKF, improving the assimilations.

As benchmarks, we obtained the operational ensemble forecasts from several operational centers. These included NCEP, ECMWF, the UK Met Office, and the CMC. Additionally, we produced another set of experimental forecasts, this time with a 30-km implementation of NOAA’s global FIM model, which utilized an icosahedral horizontal grid. The FIM ensemble was initialized from the GFS/EnKF initial conditions as well.

The experimental track forecasts from the T382 GFS with the EnKF and TCVitals observations were markedly better than the NCEP operational forecasts. They were competitive in skill with the ECMWF ensemble forecasts and substantially more accurate than the ensemble forecasts of CMC and UKMO. FIM/EnKF ensemble forecasts were not as skillful as the GFS/EnKF or ECMWF forecasts but were more skillful than those from other models. The FIM/EnKF was

the single ensemble using different dynamic cores for assimilating model vs. forecast model.

GFS/EnKF track forecast spreads were generally consistent in magnitude with the ensemble-mean errors. There were generally small correlations in the ensemble systems between the day-3 forecast spread and the day-3 ensemble-mean error, though a perfect-model analysis showed that there was a relatively modest upper bound of about 0.5 to the best-case correlations between spread and error. The ensemble forecasts did tend to stretch out preferentially in certain directions, and an analysis suggested that there was some useful information in the forecast ellipticity, though in general the ensemble forecasts should have been more isotropic in character.

Hurricane intensity forecasts remained problematic in our study. Though the experimental GFS/EnKF initialized relatively more intense vortices, they filled during the first day and produced wind-speed forecasts that were biased too low. Other forecast models did little better. It may be unrealistic to expect that global models with resolutions of 43 km and above will be able to produce realistic hurricane-like structures and associated wind forecasts, given that the radius of maximum winds is ~ 55 km on average.

We attempted to try to isolate whether the new TCVitals data, the assimilation method, or the increase in resolution were most responsible for the improvement in forecasts. Our analysis was inconclusive. In a comparison of the 3D-Var GSI assimilation systems and subsequent deterministic T382 forecasts with

the GFS, the assimilation of TC Vital observations produced only a small improvement. The T382 GFS/EnKF ensemble was used to initialize a T126 ensemble forecast; this facilitated a comparison with the operational NCEP system at the same resolution. The T126 GFS/EnKF forecasts produced dramatically better track forecasts than operational, suggesting the importance of the data assimilation with the high-resolution EnKF.

The tests shown here indicate that the combination of EnKF and the high-resolution forecast models can potentially have a large positive impact on operational tropical cyclone track forecasts. Will the results be improved even more if the forecast model resolution is refined further, so that the global model produces more realistic hurricane vortices? That is our operating hypothesis for future work. There are other aspects of the EnKF that we hope to address in future tests. It is known that if the observed position of a vortex is not within the span of the background ensemble, then the quality of the analysis may be rather poor (Chen and Snyder 2007). Perhaps the vortex relocation technology of the GSI (Liu et al. 2000) can be incorporated into the EnKF, so that the ensemble of positions does span the observed position and analysis quality is improved. Other major questions are how we can improve the treatment of the model-related uncertainty, including uncertainty effects specific to TCs, and whether regional models with EnKF initialization can be nested within the global EnKFs and provide detail on the hurricane structure and intensity not yet possible with global models.

Acknowledgments

Support for this research was provided through the NOAA HFIP program. Most of the computations of the experimental EnKF and ensemble forecasts were conducted thanks to a computational grant from the National Science Foundation to use the Texas Advanced Computing Center. We thank Bob Gall and Fred Toepfer of NOAA for helping to obtain those cycles.

References

- Aksoy, A., D. C. Dowell, C. Snyder, 2009: A multi-case comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analyses. *Mon. Wea. Rev.*, **137**, 1805-1824.
- Anderson, J. L., and Collins, N., 2007: Scalable implementations of ensemble filter algorithms for data assimilation. *J. Atmos. Ocean. Tech.*, **24**, 1452-1463.
- Arakawa, A. and W. H. Schubert, 1974: Interaction of a cumulus ensemble with the large-scale environment, part 1. *J. Atmos. Sci.*, **31**, 674-704.
- , and V. R. Lamb, 1977: Computational design of the basic dynamical processes of the UCLA general circulation model. *Methods Comput. Phys.*, **17**, 173-265.
- Barkmeijer, J., M. van Gijzen, and F. Bouttier, 1998: Singular vectors and estimates of the analysis-error covariance metric. *Quart. J. Royal Meteor. Soc.*, **124**, 1695-1713.
- , R. Buizza, and T. N. Palmer, 1999: 3D-Var Hessian singular vectors and their potential use in the ECMWF ensemble prediction system. *Quart. J. Royal Meteor. Soc.*, **125**, 2333-2351.
- Bender M. A., and I. Ginis, 2000: Real-case simulations of hurricane-ocean interaction using a high-resolution coupled model: Effects on hurricane intensity. *Mon. Wea. Rev.*, **128**, 917-946.
- , I. Ginis, R. Tuleya, B. Thomas, and T. Marchok, 2007: The Operational GFDL Coupled Hurricane-Ocean Prediction System and a Summary of Its Performance. *Mon. Wea. Rev.*, **135**, 3965-3989.

- Bishop, C.H., B.J. Etherton, and S.J. Majumdar, 2001: Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. *Mon. Wea. Rev.*, **129**, 420–436.
- Bleck, R., S. G. Benjamin, J. L. Lee, A. E. MacDonald, 2010: On the use of an arbitrary Lagrangian-Eulerian vertical coordinate in global atmospheric modeling. *Mon. Wea. Rev.*, in press.
- Bowler, N., A. Arribas, K. R. Mylne, K. B. Robertson, and S. E. Beare, 2008: The MOGREPS short-range ensemble prediction system. *Quart. J. Royal Meteor. Soc.*, **134**, 703-722.
- , -----, S. E. Beare, K. R. Mylne, and G. J. Shutts, 2009: The local ETKF and SKEB: Upgrades to the MOGREPS short-range ensemble prediction system. *Quart. J. Royal Meteor. Soc.*, **135**, 767-776.
- Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He, 2010a: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: description of configurations and results from idealized experiments. *Mon. Wea. Rev.*, in press. Available from Mark.Buehner@ec.gc.ca.
- , -----, -----, -----, and -----, 2010b: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: results from 1-month experiments with real observations. *Mon. Wea. Rev.*, in press. Available from Mark.Buehner@ec.gc.ca.
- Buizza, R., and T. N. Palmer, 1995: The singular-vector structure of the atmospheric

- global circulation. *J. Atmos. Sci.*, **52**, 1434–1456.
- , 1997: Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system. *Mon. Wea. Rev.*, **125**, 99–119.
- , J.-R. Bidlot, N. Wedi, M. Fuentes, M. Hamrud, G. Holt, and F. Vitart, 2007: The new ECMWF VAREPS (Variable-Resolution Ensemble Prediction System). *Quart. J. Royal Meteor. Soc.*, **133**, 681–695.
- Charron, M., G. Pellerin, L. Spacek, P. L. Houtekamer, N. Gagnon, H. L. Mitchell, and L. Michelin, 2010: Towards random sampling of model error in the Canadian ensemble prediction system. *Mon. Wea. Rev.*, in review. Available from Martin.Charron@ec.gc.ca.
- Chen, Y., and C. Snyder, 2007: Assimilating vortex position with an ensemble Kalman filter. *Mon. Wea. Rev.*, **135**, 1828–1845.
- Chou, M.-D., 1992: A solar radiation model for use in climate studies. *J. Atmos. Sci.*, **49**, 762–772.
- , and K. T. Lee, 1996: Parameterizations for the absorption of solar radiation by water vapor and ozone. *J. Atmos. Sci.*, **53**, 1204–1208.
- Chou, M.-D., M. J. Suarez, C. H. Ho, M. M. H. Yan, and K. T. Lee, 1998: Parameterizations for cloud overlapping and shortwave single scattering properties for use in general circulation and cloud ensemble models. *J. Climate*, **11**, 202–214.

- Compo, G.P., J.S. Whitaker, and P.D. Sardeshmukh, 2006: Feasibility of a 100 year reanalysis using only surface pressure data. *Bull. Amer. Met. Soc.*, **87**, 175-190.
- Courtier, P., J.-N. Thepaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Royal Meteor. Soc.*, **120**, 1367-1387.
- Davis, C., W. Wang, S.S. Chen, Y. Chen, K. Corbosiero, M. DeMaria, J. Dudhia, G. Holland, J. Klemp, J. Michalakes, H. Reeves, R. Rotunno, C. Snyder, and Q. Xiao, 2008: Prediction of Landfalling Hurricanes with the Advanced Hurricane WRF Model. *Mon. Wea. Rev.*, **136**, 1990–2005.
- Ehrendorfer, M., 2007: A review of issues in ensemble Kalman filtering. *Meteor. Zeitschrift*, **16**, 795-818.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99(C5)**, 10143-10162.
- , 2009: *Data Assimilation, the Ensemble Kalman Filter (2nd Ed)*. Springer Press, 307 pp.
- Fiorino, M., and R. L. Elsberry, 1989a: Some aspects of vortical structure related to tropical cyclone motion. *J. Atmos. Sci.*, **46**, 975-990.
- Fiorino, M., and R. L. Elsberry, 1989b: Contributions to tropical cyclone motion by small, medium, and large scales in the initial vortex. *Mon. Wea. Rev.*, **117**, 721-727.
- Fiorino, M., 2009: Record-setting performance of the ECMWF IFS in medium-range

- tropical cyclone track prediction. *ECMWF Newsletter*, **188**, 20-27.
- Gentry, M. S., and G. M. Lackmann, 2010: Sensitivity of simulated tropical cyclone structure and intensity to horizontal resolution. *Mon. Wea. Rev.*, **138**, in press.
- Global Climate and Weather Modeling Branch, 2003: The GFS Atmospheric Model. NCEP Office Note 442, 14 pp. Available at <http://tinyurl.com/yc7bkre>
- Grasso, L. D., 2000: The difference between grid spacing and resolution and their application to numerical modeling. *Bull. Amer. Meteor. Soc.*, **81**, 579–580.
- Grell, G. A., 1993: Prognostic evaluation of assumptions used by cumulus parameterizations. *Mon. Wea. Rev.*, **121**, 764–787.
- Guard, C. P., L. E. Carr, F. H. Wells, R. A. Jeffries, N. D. Gural, D. K. Edson, 1992: Joint Typhoon Warning Center and the Challenges of Multibasin Tropical Cyclone Forecasting. *Wea. Forecasting*, **7**, 328–352.
- Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155-167.
- , 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550-560.
- , 2006: Ensemble-based atmospheric data assimilation. Chapter 6 in *Predictability of Weather and Climate*, T. N. Palmer and R. Hagedorn, Eds. Cambridge Press, 124-156.
- Hong, S.-Y. and H.-L. Pan, 1996: Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Mon. Wea. Rev.*, **124**, 2322–2339.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796-811.

-----, -----, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric data assimilation with an ensemble Kalman filter: results with real observations. *Mon. Wea. Rev.*, **133**, 604-620.

-----, -----, 2005: Ensemble Kalman filtering. *Quart. J. Roy. Meteor. Soc.*, **131**, 3269–3289.

-----, -----, and X. Deng, 2009: Model error representation in an operational ensemble Kalman filter. *Mon. Wea. Rev.*, **137**, 2126-2143.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, **230**, 112-126. <http://dx.doi.org/10.1016/j.physd.2006.11.008>

Kimball, S.K., and M.S. Mulekar, 2004: A 15-year climatology of North Atlantic tropical cyclones. Part I: size parameters. *J. Climate*, **17**, 3555–3575.

Kleist, D.T., D.F. Parrish, J.C. Derber, R. Treadon, W.S. Wu, and S. Lord, 2009: Introduction of the GSI into the NCEP Global Data Assimilation System. *Wea. Forecasting*, **24**, 1691–1705.

Krishnamurti, T., D. Oosterhof, and N. Dignon, 1989: Hurricane prediction with a high resolution global model. *Mon. Wea. Rev.*, **117**, 631–669.

-----, S. Pattnaik, L. Stefanova, T. S. V. Vijaya Kumar, B. P. Mackey, A. J. O'Shay, and R. J. Pasch, 2005: The hurricane intensity issue. *Mon. Wea. Rev.*, **133**, 1886–1912.

- Lam, C. C., 2001: Performance of the ECMWF model in forecasting the tracks of tropical cyclones in the South China Sea and parts of the western North Pacific. *Meteor. Appl.*, **8**, 339-344.
- Le-Dimet, F.-X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations. Theoretical aspects. *Tellus*, **38A**, 97-110.
- Lee, J. L., and A. E. MacDonald, 2009: A Finite-Volume Icosahedral Shallow-Water Model on a Local Coordinate. *Mon. Wea. Rev.*, **137**, 1422-1437.
- Leutbecher, M., 2005: On ensemble prediction using singular vectors started from forecasts. *Mon. Wea. Rev.*, **133**, 3038-3046.
- Li, X., and Z. Pu, 2008: Sensitivity of Numerical Simulation of Early Rapid Intensification of Hurricane Emily (2005) to Cloud Microphysical and Planetary Boundary Layer Parameterizations. *Mon. Wea. Rev.*, **136**, 4819-4838.
- Liu, Q., T. Marchok, H.-L. Pan, M. Bender, and S. J. Lord, 2000: Improvements in hurricane initialization and forecasting at NCEP with global and regional models. *NWS Tech Procedures Bulletin 472*, available at <http://www.nws.noaa.gov/om/tpb/472.htm>.
- Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP – a comparison with 4D-Var. *Quart. J. Royal Meteor. Soc.*, **129**, 3183-3203.
- Majumdar, S., and P. M. Finocchio, 2010: On the ability of global ensemble prediction systems to predict tropical cyclone track probabilities. *Wea. Forecasting*, in press. Available from smajumdar@rsmas.miami.edu.

- Meng, Z., and F. Zhang, 2008a: Test of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part III: Comparison with 3DVar in a real-data case study. *Mon. Wea. Rev.*, **136**, 522-540.
- , and -----, 2008b: Test of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part IV: Comparison with 3DVar in a month-long experiment. *Mon. Wea. Rev.*, **136**, 3671-3682.
- Miyoshi, T., and S. Yamane, 2007: Local Ensemble Transform Kalman Filtering with an AGCM at a T159/L48 Resolution. *Mon. Wea. Rev.*, **135**, 3841-3861.
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16663-16682.
- Molteni F., R. Buizza R., T. N. Palmer, and T. Petroliaxis, 1996. The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Royal Meteor. Soc.*, **122**, 73-119.
- Palmer, T. N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. J. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic parameterization and model uncertainty. *ECMWF Technical Memorandum 589*. Available from <http://www.ecmwf.int/publications/library/do/references/list/14>.
- Pan, H.-L. and W.-S. Wu, 1995: Implementing a mass flux convection parameterization package for the NMC Medium-Range Forecast Model. NMC Office Note, No. 409, 40pp. [Available from NCEP, 5200 Auth Road, Washington, DC 20233]

- Rabier, F., H. Jarvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation: I: experimental results with simplified physics. *Quart. J. Royal Meteor. Soc.*, **126**, 1143-1170.
- Rappaport, E.N., J.L. Franklin, L.A. Avila, S.R. Baig, J.L. Beven, E.S. Blake, C.A. Burr, J.G. Jiing, C.A. Juckins, R.D. Knabb, C.W. Landsea, M. Mainelli, M. Mayfield, C.J. McAdie, R.J. Pasch, C. Sisko, S.R. Stewart, and A.N. Tribble, 2009: Advances and Challenges at the National Hurricane Center. *Wea. Forecasting*, **24**, 395–419.
- Rawlins, F., S. P. Ballard, K. J. Bovis, A. M. Clayton, D. Li, G. W. Inverarity, A. C. Lorenc, and T. J. Payne: The Met Office global four-dimensional variational data assimilation scheme. *Quart. J. Royal Meteor. Soc.*, **133**, 347-362.
- Rogers R., and coauthors, 2006: The intensity forecasting experiment: A NOAA multiyear field program for improving tropical cyclone intensity forecasts. *Bull. Amer. Meteor. Soc.*, **87**, 1523–1537.
- Shutts G., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Royal Meteor. Soc.*, **131**, 3079–3102.
- Skamarock, W. C., 2004: Evaluating Mesoscale NWP Models Using Kinetic Energy Spectra. *Mon. Wea. Rev.*, **132**, 3019–3032.
- Strang, G., 1988: *Linear algebra and its applications*. Harcourt Brace Jovanovich, 505 pp.
- Sundqvist, H., E. Berge, and J. E. Kristjansson, 1989: Condensation and cloud studies with mesoscale numerical weather prediction model. *Mon. Wea. Rev.*, **117**,

1641–1757.

Szunyogh, I., E. J. Kostelich, G. Gyarmati, E. Kalnay, B. R. Hunt, E. Ott, E. Satterfield, and J. A. Yorke, 2008: A local ensemble transform Kalman filter data assimilation system for the NCEP global model. *Tellus*, **60A**, 113–130.

Tiedtke, M., 1983: The sensitivity of the time-mean large-scale flow to cumulus convection in the ECMWF model. *ECMWF Workshop on Convection in Large-Scale Models*, 297–316.

Torn, R. D., and G. J. Hakim, 2008: Performance characteristics of a pseudo-operational ensemble Kalman filter. *Mon. Wea. Rev.*, **136**, 3947–3963.

-----, and -----, 2009: Ensemble Data Assimilation applied to RAINEX observations of Hurricane Katrina (2005). *Mon. Wea. Rev.*, **137**, 2817–2829.

Troen, I. and L. Mahrt, 1986: A simple model of the atmospheric boundary layer; sensitivity to surface evaporation. *Bound. Layer Meteor.*, **37**, 129–148.

Wang, X., and C.H. Bishop, 2003: A Comparison of Breeding and Ensemble Transform Kalman Filter Ensemble Forecast Schemes. *J. Atmos. Sci.*, **60**, 1140–1158.

-----, D. M. Barker, C. Snyder, and T. M. Hamill, 2008: A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part II: real observation experiments. *Mon. Wea. Rev.*, **136**, 5132–5147.

Wei, M. and Z. Toth, R. Wobus, Y. Zhu, 2008: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system. *Tellus*, **60A**, 62–79.

Whitaker, J. S., and A. F. Lough, 1998: The relationship between ensemble spread

- and ensemble mean skill. *Mon. Wea. Rev.*, **126**, 3292-3302.
- , and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913-1924.
- , G.P. Compo, X. Wei and T.M. Hamill, 2004: Reanalysis without radiosondes using ensemble data assimilation. *Mon. Wea. Rev.*, **132**, 1190-1200.
- , T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP global forecast system. *Mon. Wea. Rev.*, **136**, 463-482.
- Wu, W.S., R.J. Purser, and D.F. Parrish, 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.*, **130**, 2905-2916.
- Zhang, D.-L., and X. Wang, 2003: Dependence of hurricane intensity and structures on vertical resolution and time-step size. *Adv. In Atm. Sci.*, **20**, 711-725.
- , Y. Weng, Z. Meng, J. A. Sippel and C. H. Bishop, 2009: Cloud-resolving hurricane initialization and prediction through assimilation of Doppler radar observations with an ensemble Kalman Filter: Humberto (2007). *Mon. Wea. Rev.*, **137**, 2105-2125.
- Zhao, Q. Y. and F. H. Carr, 1997: A prognostic cloud scheme for operational NWP models. *Mon. Wea. Rev.*, **125**, 1931-1953.

LIST OF FIGURE CAPTIONS

Fig. 1: Illustration of the potential of an EnKF for making meteorologically reasonable increments. (a) Mean background forecast of 850-hPa winds. Assume an observation 10 ms^{-1} greater than the background is available at the black dot. (b) Wind analysis increment from the assimilation of the single observation when using an EnKF. (c) Wind analysis increment when using an average of then ensemble covariances over all locations around the hurricane, a surrogate for the increment that may be expected from a 3-dimensional variational data assimilation system.

Fig. 2: Tropical cyclones in the Atlantic basin during the period of the experiment. Storms are not counted if they are not at least tropical depression in strength.⁴

Fig. 3: As in Fig. 2, but for the eastern-Pacific basin. The one un-named track (northwest of Kevin) never reached tropical-storm strength.

Fig. 4: As in Fig. 2, but for the western-Pacific basin.

Fig. 5: Homogeneous comparison of global average track forecast errors and average spread between the experimental GFS/EnKF and (a) the NCEP operational ensemble system, (b) the ECMWF operational ensemble system, (c) the CMC operational ensemble system, (d) the UK Met Office operational ensemble system, and (e) the experimental FIM/EnKF ensemble system. Numbers in parentheses up top on each panel indicate the sample size at a particular forecast lead, i.e., the number of matched paired forecasts between the GFS and the model in question. Dashed lines indicate spread, solid lines indicate error. Error bars indicate the 5th and 95th percentiles of a resampled block bootstrap distribution following Hamill

(1999).

Fig. 6: Ensemble forecasts of hurricane Bill cyclone positions from global ensembles initialized at 0000 UTC 19 Aug 2009. Observed track in black, with 0000 UTC positions denoted by the “*” sign. Black colors indicate day-0 positions; orange = day+1 positions; purple = day+2; red = day+3; yellow-green = day+4; blue = day+5 positions. Small dots indicate ensemble member positions, large dots indicate ensemble mean. Ellipses generated from bivariate normal distribution fitted to the forecast positions; contour encloses 90% of the probability.

Fig. 7: As in Fig. 6, but for ensemble forecasts of typhoon Morakot cyclone positions from global ensembles initialized at 0000 UTC 5 August 2009. FIM data was missing on this day.

Fig. 8: As in Fig. 6, but for ensemble forecasts of hurricane Jimena cyclone positions from global ensembles initialized at 0000 UTC 30 August 2009. FIM/EnKF data is missing.

Fig. 9: 3-day forecast spread of cyclone position (calculated according to eq. 2) plotted against the absolute error of the ensemble-mean position for each of the tested ensemble systems. Spearman rank correlation r and sample size n are also noted. Data is non-homogeneous.

Fig. 10: As in Fig. 9, but a perfect-model experiment where the first tracked forecast member is used as a synthetic verification.

Fig. 11: Schematic illustration of the decomposition of spread and errors into the eigenspace of the sample forecast-error covariance. Small black dots denote ensemble forecast member cyclone positions; the large black dot denotes the

ensemble-mean position. Contour indicates shape of bi-variate normal distribution fit to the data. Black arrows S_1 and S_2 represent the eigenvectors of the covariance matrix; for illustration purposes here, the eigenvectors are scaled by the square root of the eigenvalue. Red dot indicates the position of the observed cyclone; red arrow indicates the ensemble-mean error vector. Dashed lines indicate the projections of the error vector \mathbf{E} onto the eigenvectors $\mathbf{E} \cdot \mathbf{S}_1$ and $\mathbf{E} \cdot \mathbf{S}_2$.

Fig. 12: Analysis of the sample-average ellipticity of the forecast ensembles and the relative correspondence of forecast error with the ellipticity. Solid lines indicate the average square roots of the eigenvalues of the forecast-error covariance matrix, blue for the leading eigenvector, red for the trailing eigenvector. Dashed lines indicate the average projection of the forecast error onto eigenvectors, blue dashed line the projection onto the leading eigenvector, red dashed line the projection onto the trailing eigenvector. Data here is non-homogeneous.

Fig. 13: Wind speed bias (average ensemble-mean forecast minus observed) for the forecast models used in this study (UKMO did not report wind speed). Error bars represent the 5th and 95th percentiles of a normal distribution fit to the available samples. Data here is non-homogeneous.

Fig. 14: Scatterplots of analyzed maximum wind speed (abscissa) vs. 1-day forecast wind speed (ordinate) for (a) GFS/EnKF deterministic forecast from ensemble-mean initial condition, and (b) GSI-parallel deterministic forecast. (c) Observed wind speed from TC Vitals (abscissa) vs. analyzed wind speed (ordinate) for GFS/EnKF, and (d) GSI-Parallel.

Fig. 15: Background forecast (black contours; interval 1 hPa) of sea-level pressure, and mean analysis increment (dashed lines and colors) for hurricane Ike at 0000 UTC 4 September 2008. (a) Parallel run of T382 GSI, and (b) T382 GFS.

Fig. 16: (a) Homogeneous track error comparison for deterministic forecasts initialized from the GFS/EnKF ensemble-mean initial condition (blue), the GSI-parallel (yellow-green) and the GSI operational (red). Error bars represent the 95th percentiles of a paired block bootstrap distribution between GSI operational and GSI parallel. (b) Homogeneous track error comparison for deterministic forecasts initialized from the GFS/EnKF ensemble-mean initial condition (blue) vs. the mean of the ensemble-forecasts positions from the experimental GFS/EnKF (red).

Fig. 17: Track error and spread, as in Fig. 5, but for (a) comparison of experimental GFS/EnKF with ensemble forecasts conducted at T382L64 resolution vs. experimental GFS/EnKF with ensemble forecasts conducted at the resolution of the 2009 operational system T126L28 (though initial conditions from T382L64 EnKF). (b) Comparison of operational T126L28 NCEP ensemble forecast system track errors to the experimental GFS/EnKF at the same resolution.

Fig. 18. Anomaly correlation of tropical (equatorward of 20 degrees) 250-hPa zonal wind from deterministic 72-h T382L64 GFS forecasts from the operational GSI analysis (blue) and the experimental GFS/EnKF analysis (red).

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
					<i>Jul 31: F</i>	Aug 1: F
2: F,E	3: F	4: F	5: F	6: F,U	7: F	8: F
9: F	10: F,G	11: F	12: F	13	14	15
16	17	18	19	20	21	22
23: U	24: F	25	26: G	27: F	28: F	29: F
30	31	<i>Sep 1: G</i>	2	3	<i>4: F</i>	5
6	<i>7: F</i>	8	9	10	11	12
13	14	<i>15: C</i>	16	17	18	19
20	21	22	<i>23: F,E</i>	<i>24: F</i>	25	<i>26: E</i>
<i>27: E,U</i>	<i>28: U</i>					

Table 1: Availability of 0000 UTC global ensemble forecast data between 31 July 2009 and 28 September 2009. For a particular date, “F” indicates that FIM ensemble data was unavailable for this initial time; E indicates that ECMWF ensemble was unavailable; U indicates UKMO; C indicates CMC; N indicates NCEP, and G indicates experimental GEFS/EnKF.

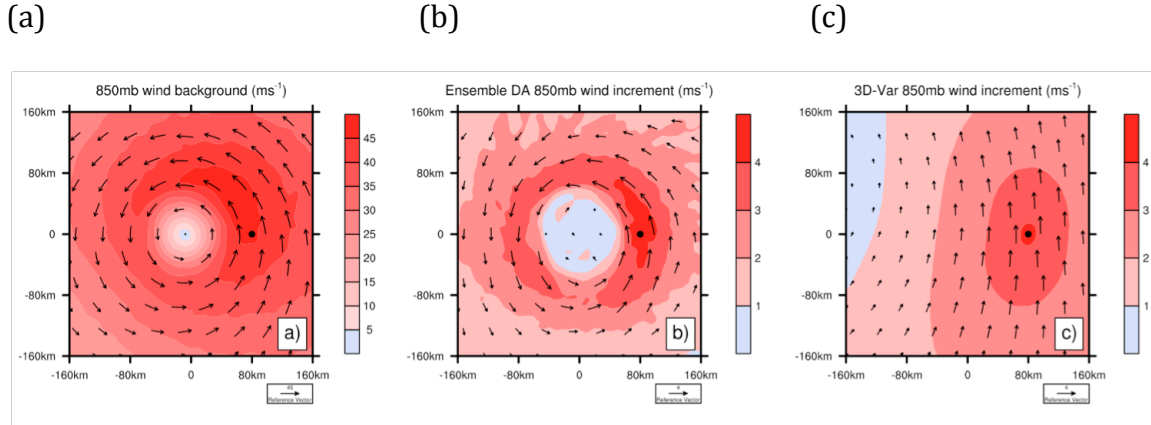


Fig. 1: Illustration of the potential of an EnKF for making meteorologically reasonable increments. (a) Mean background forecast of 850-hPa winds. Assume an observation 10 ms⁻¹ greater than the background is available at the black dot. (b) Wind analysis increment from the assimilation of the single observation when using an EnKF. (c) Wind analysis increment when using an average of the ensemble covariances over all locations around the hurricane, a surrogate for the increment that may be expected from a 3-dimensional variational data assimilation system.

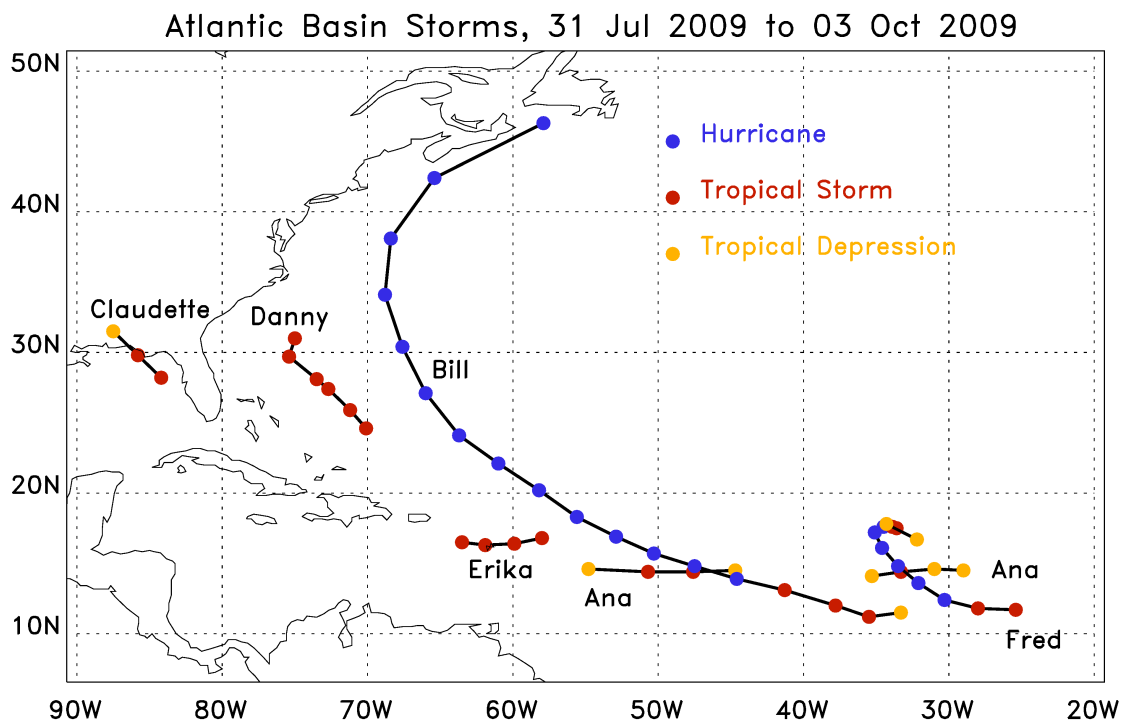


Fig. 2: Tropical cyclones in the Atlantic basin during the period of the experiment. Storms are not counted if they are not at least tropical depression in strength.

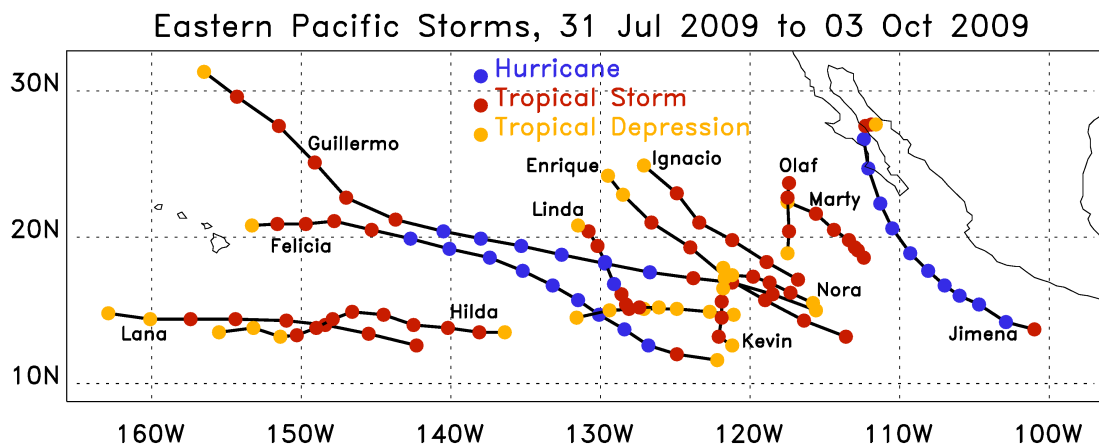


Fig. 3: As in Fig. 2, but for the eastern-Pacific basin. The one un-named track (northwest of Kevin) never reached tropical-storm strength.

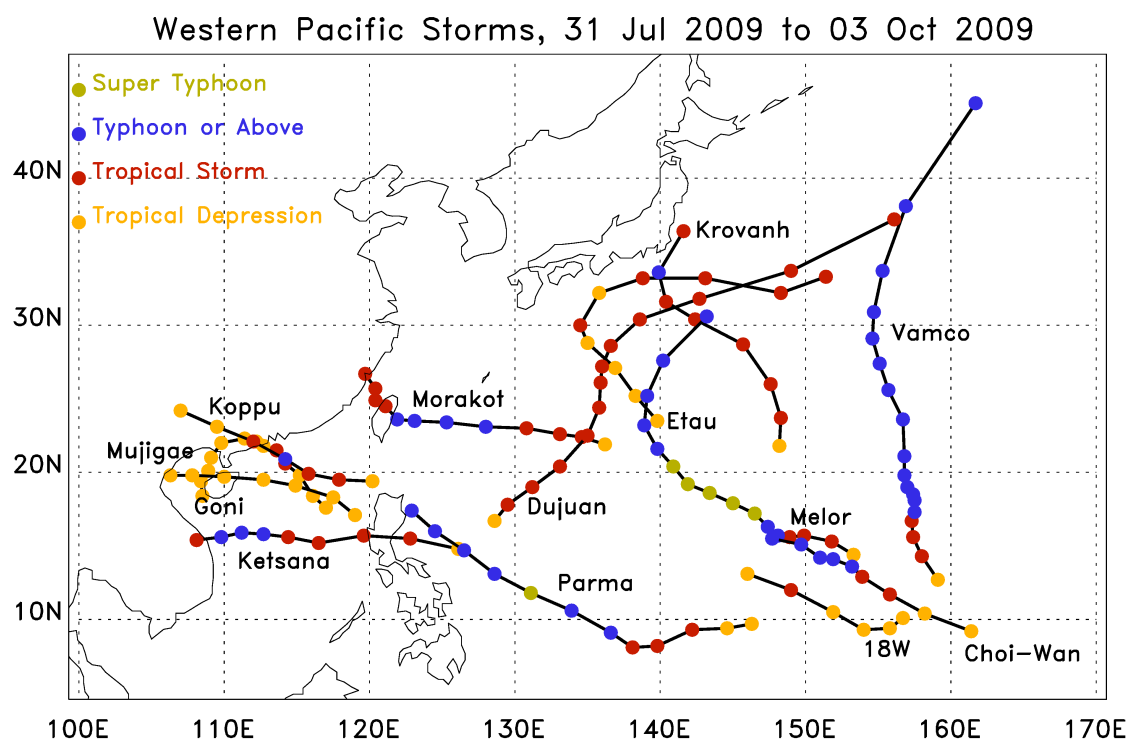


Fig. 4: As in Fig. 2, but for the western-Pacific basin.

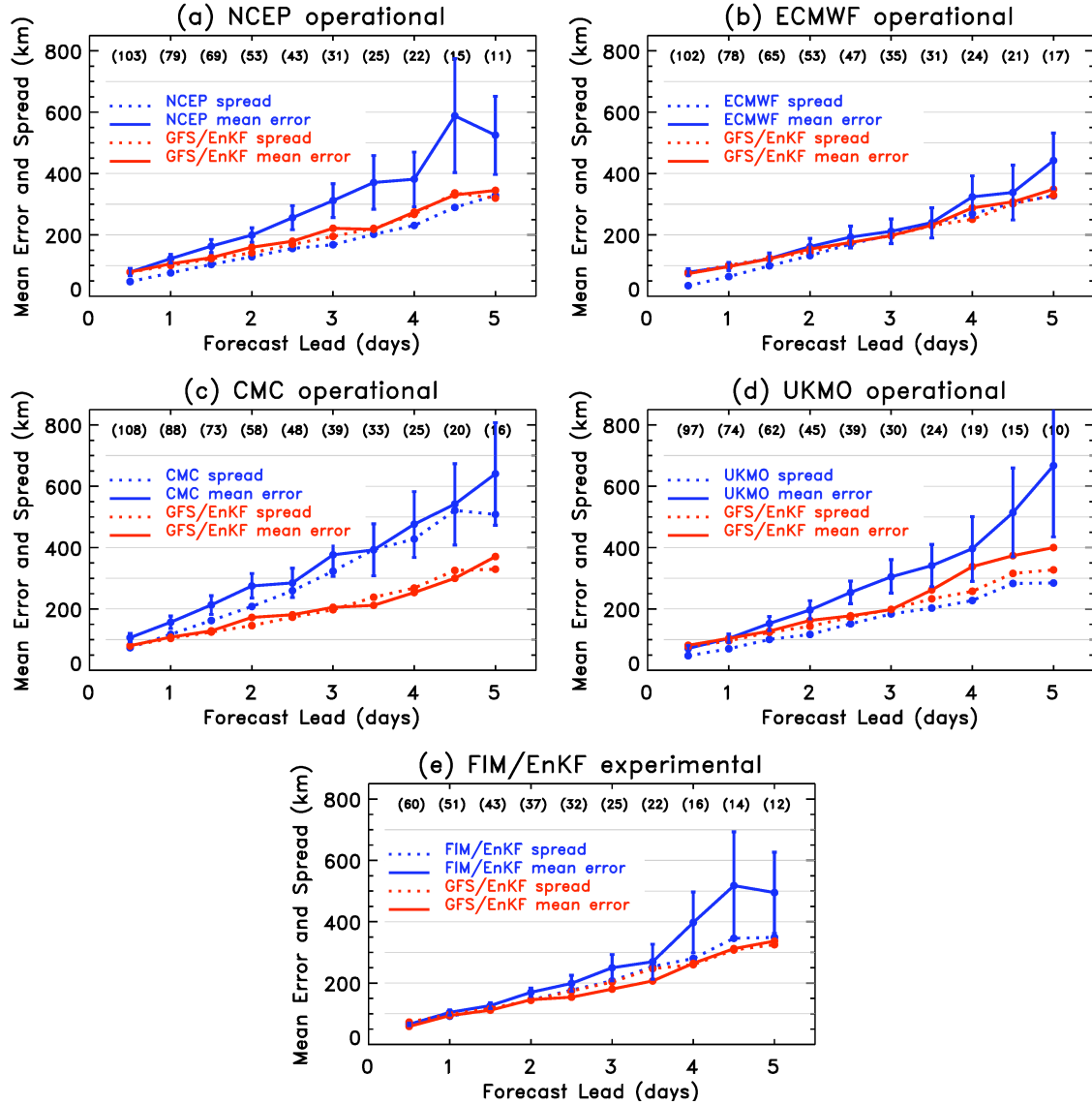


Fig. 5: Homogeneous comparison of global average track forecast errors and average spread between the experimental GFS/EnKF and (a) the NCEP operational ensemble system, (b) the ECMWF operational ensemble system, (c) the CMC operational ensemble system, (d) the UK Met Office operational ensemble system, and (e) the experimental FIM/EnKF ensemble system. Numbers in parentheses up top on each panel indicate the sample size at a particular forecast lead, i.e., the number of matched paired forecasts between the GFS and the model in question. Dashed lines indicate spread, solid lines indicate error. Error bars indicate the 5th and 95th percentiles of a resampled block bootstrap distribution following Hamill (1999).

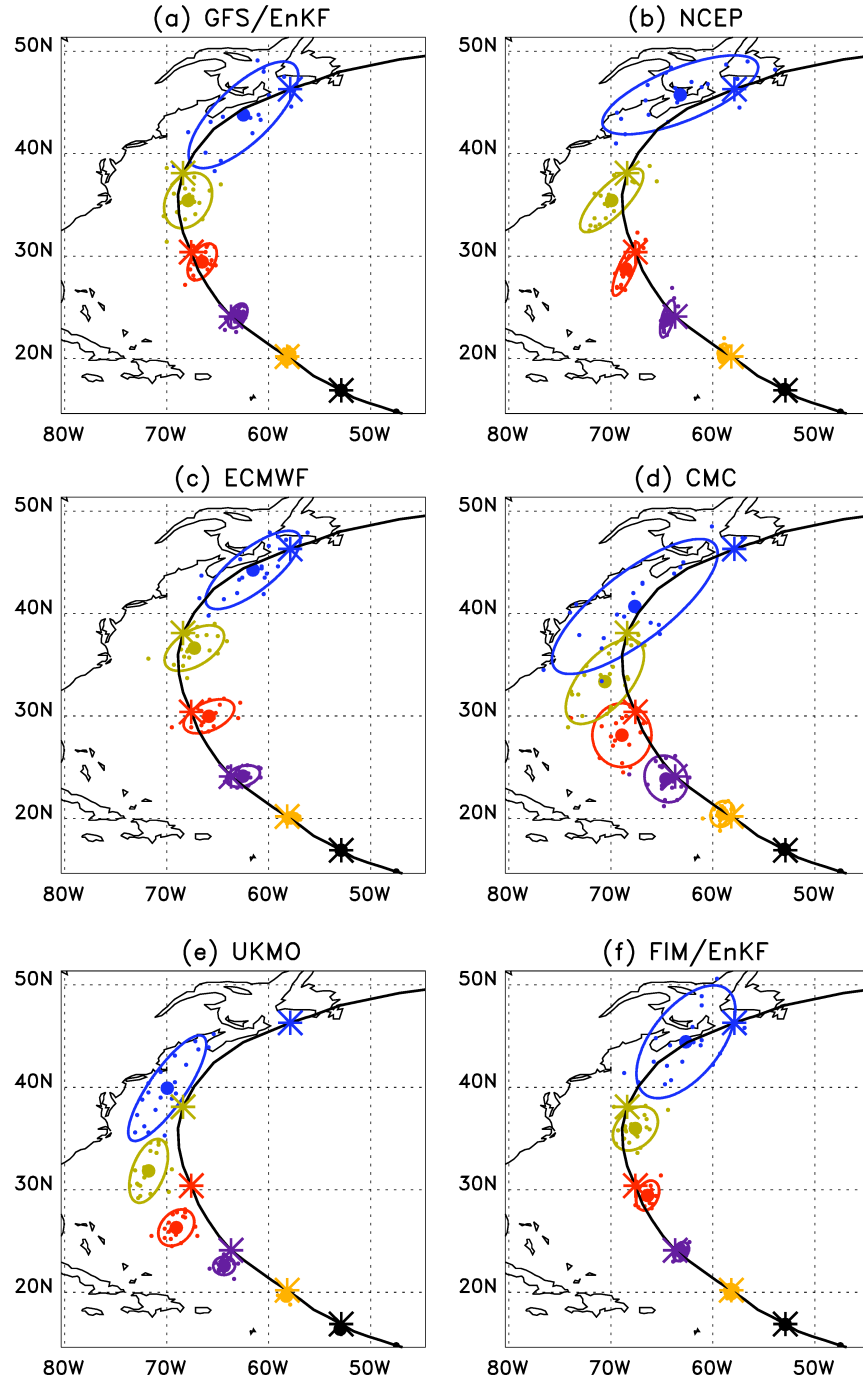


Fig. 6: Ensemble forecasts of hurricane Bill cyclone positions from global ensembles initialized at 0000 UTC 19 Aug 2009. Observed track in black, with 0000 UTC positions denoted by the “*” sign. Black colors indicate day-0 positions; orange = day+1 positions; purple = day+2; red = day+3; yellow-green = day+4; blue = day+5 positions. Small dots indicate ensemble member positions, large dots indicate ensemble mean. Ellipses generated from bivariate normal distribution fitted to the forecast positions; contour encloses 90% of the probability.

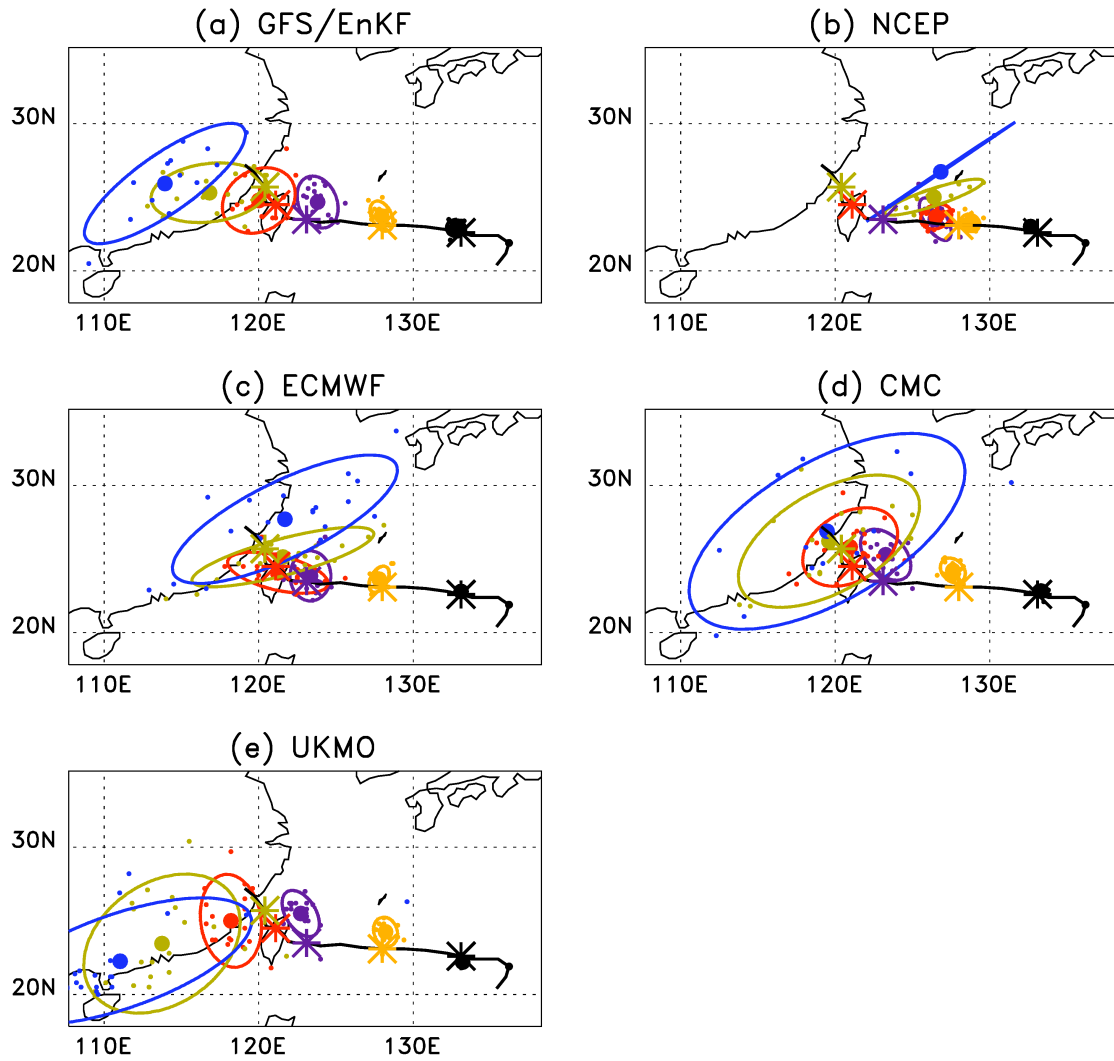


Fig. 7: As in Fig. 6, but for ensemble forecasts of typhoon Morakot cyclone positions from global ensembles initialized at 0000 UTC 5 August 2009. FIM data was missing on this day.

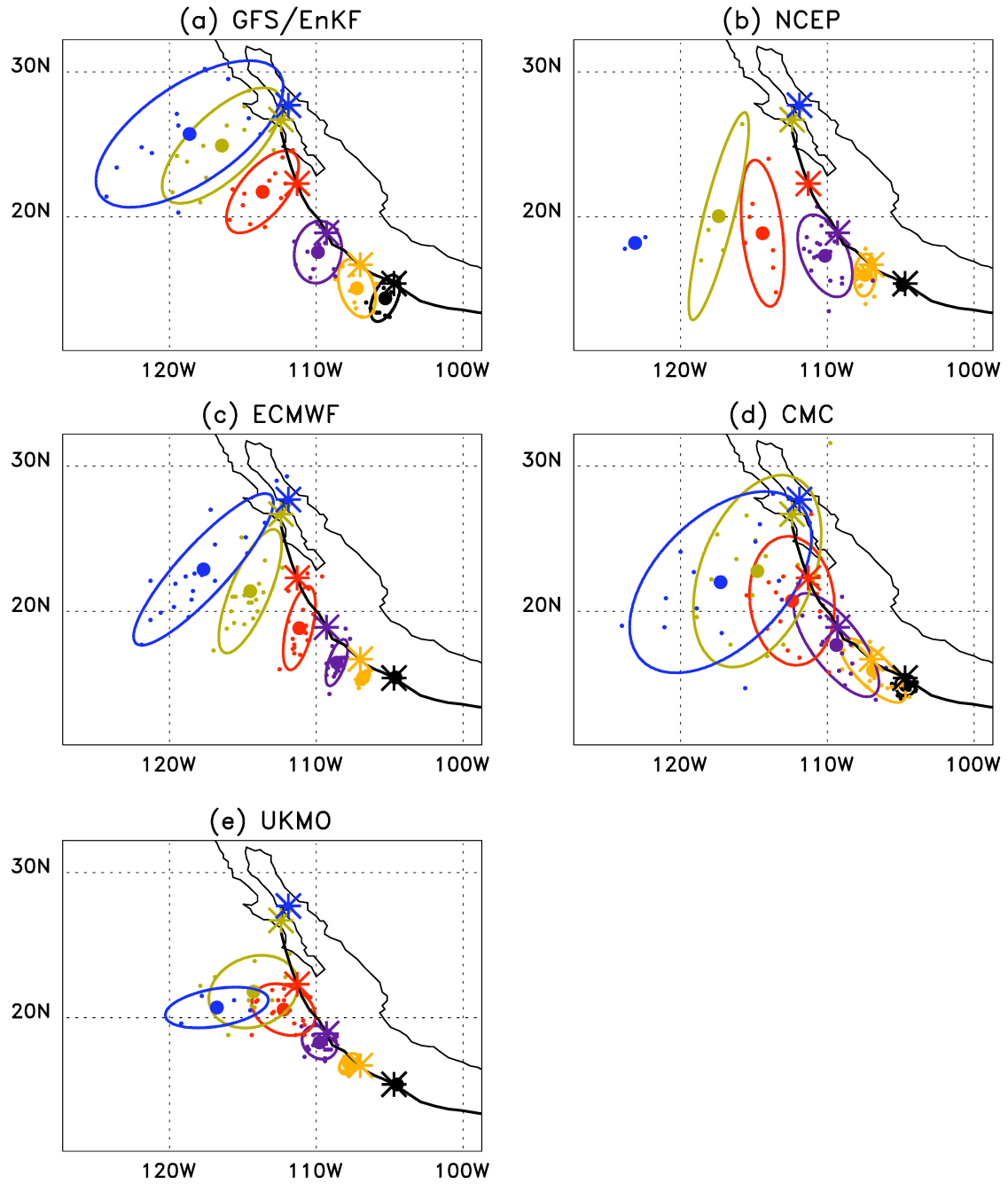


Fig. 8: As in Fig. 6, but for ensemble forecasts of hurricane Jimena cyclone positions from global ensembles initialized at 0000 UTC 30 August 2009. FIM/EnKF data is missing.

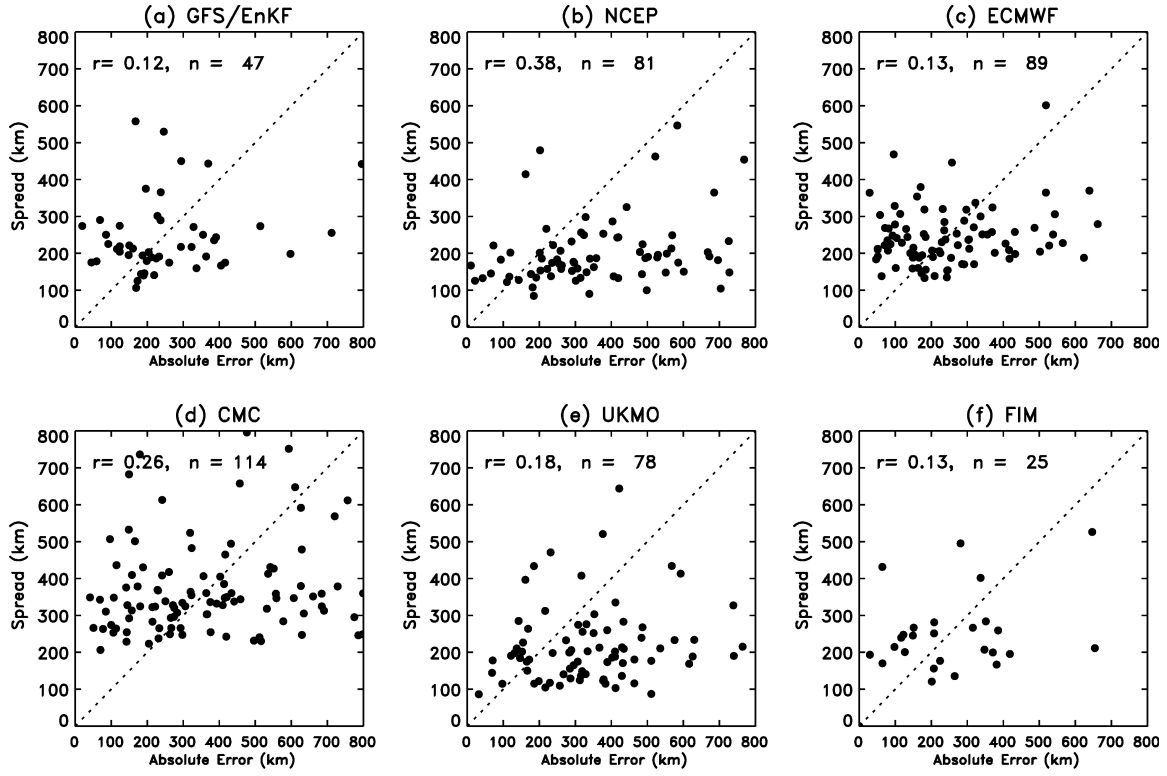


Fig. 9: 3-day forecast spread of cyclone position (calculated according to eq. 2) plotted against the absolute error of the ensemble-mean position for each of the tested ensemble systems. Spearman rank correlation r and sample size n are also noted. Data is non-homogeneous.

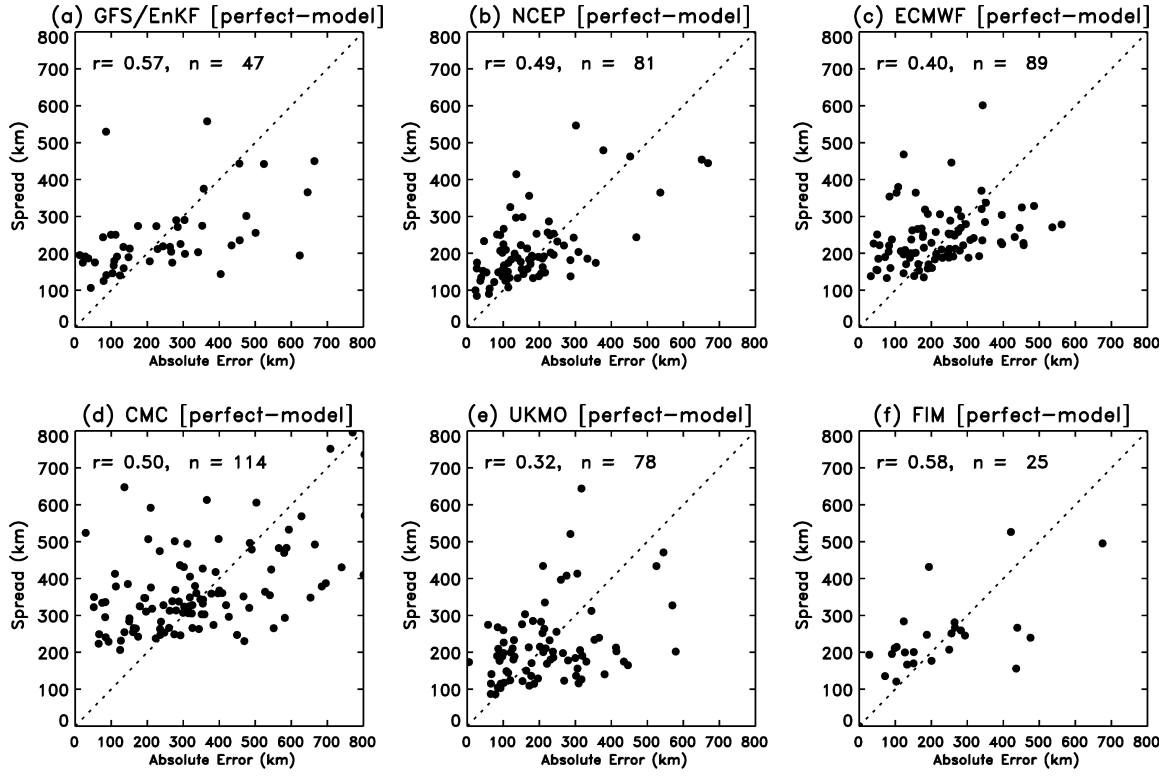


Fig. 10: As in Fig. 9, but a perfect-model experiment where the first tracked forecast member is used as a synthetic verification.

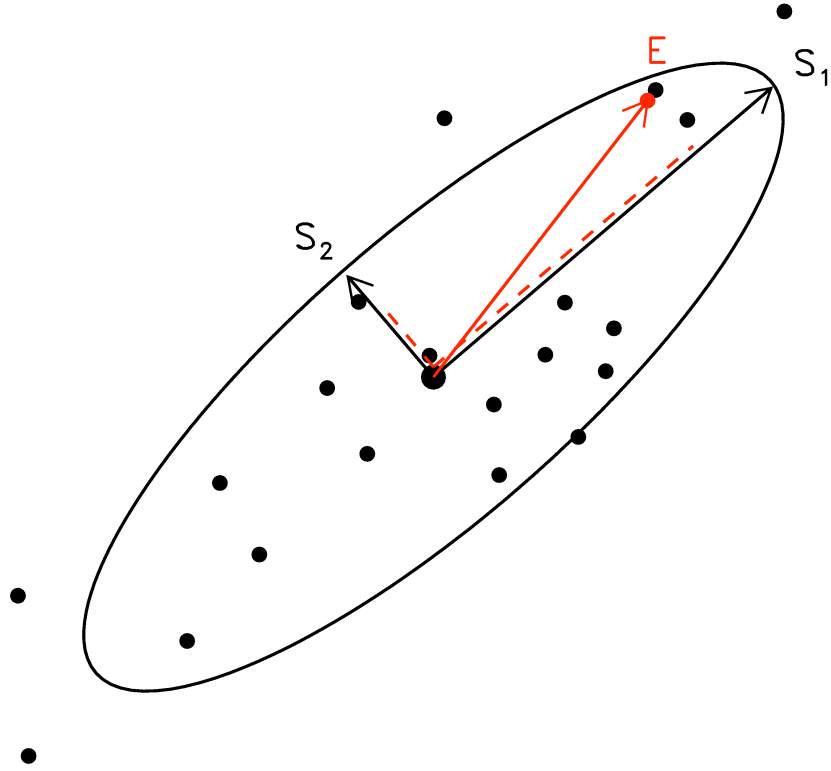


Fig. 11: Schematic illustration of the decomposition of spread and errors into the eigenspace of the sample forecast-error covariance. Small black dots denote ensemble forecast member cyclone positions; the large black dot denotes the ensemble-mean position. Contour indicates shape of bi-variate normal distribution fit to the data. Black arrows S_1 and S_2 represent the eigenvectors of the covariance matrix; for illustration purposes here, the eigenvectors are scaled by the square root of the eigenvalue. Red dot indicates the position of the observed cyclone; red arrow indicates the ensemble-mean error vector. Dashed lines indicate the projections of the error vector \mathbf{E} onto the eigenvectors $\mathbf{E} \cdot \mathbf{S}_1$ and $\mathbf{E} \cdot \mathbf{S}_2$.

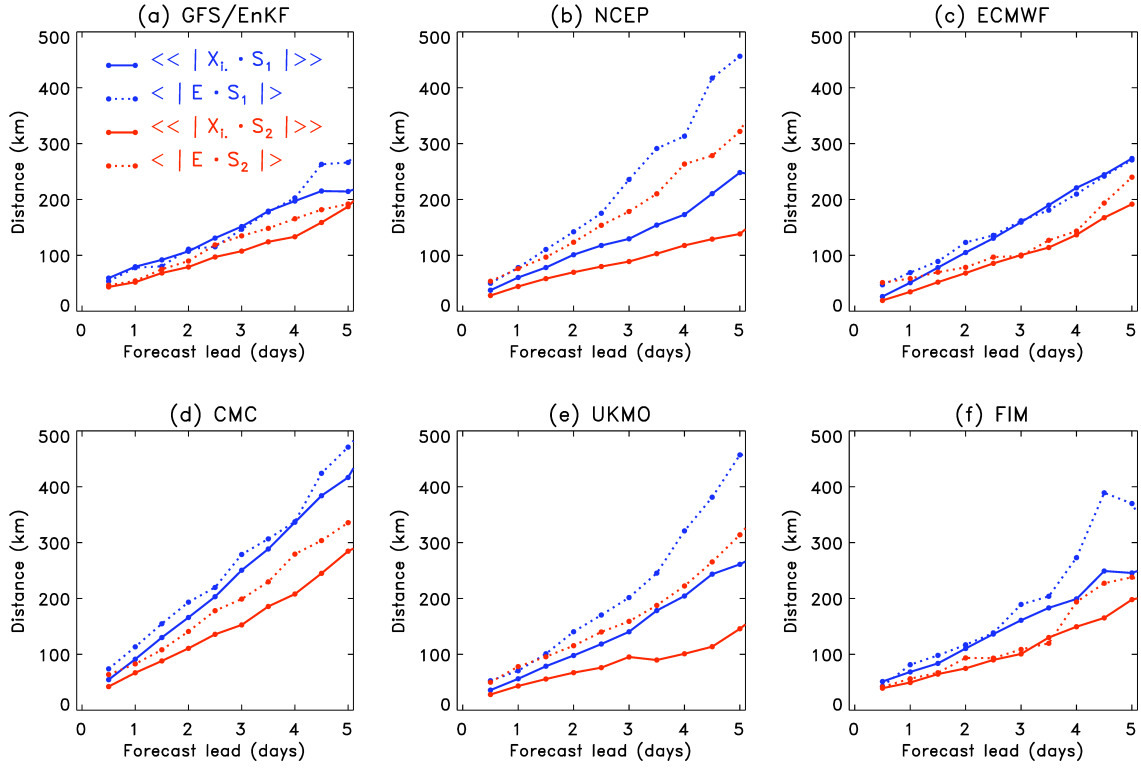


Fig. 12: Analysis of the sample-average ellipticity of the forecast ensembles and the relative correspondence of forecast error with the ellipticity. Solid lines indicate the average square roots of the eigenvalues of the forecast-error covariance matrix, blue for the leading eigenvector, red for the trailing eigenvector. Dashed lines indicate the average projection of the forecast error onto eigenvectors, blue dashed line the projection onto the leading eigenvector, red dashed line the projection onto the trailing eigenvector. Data here is non-homogeneous.

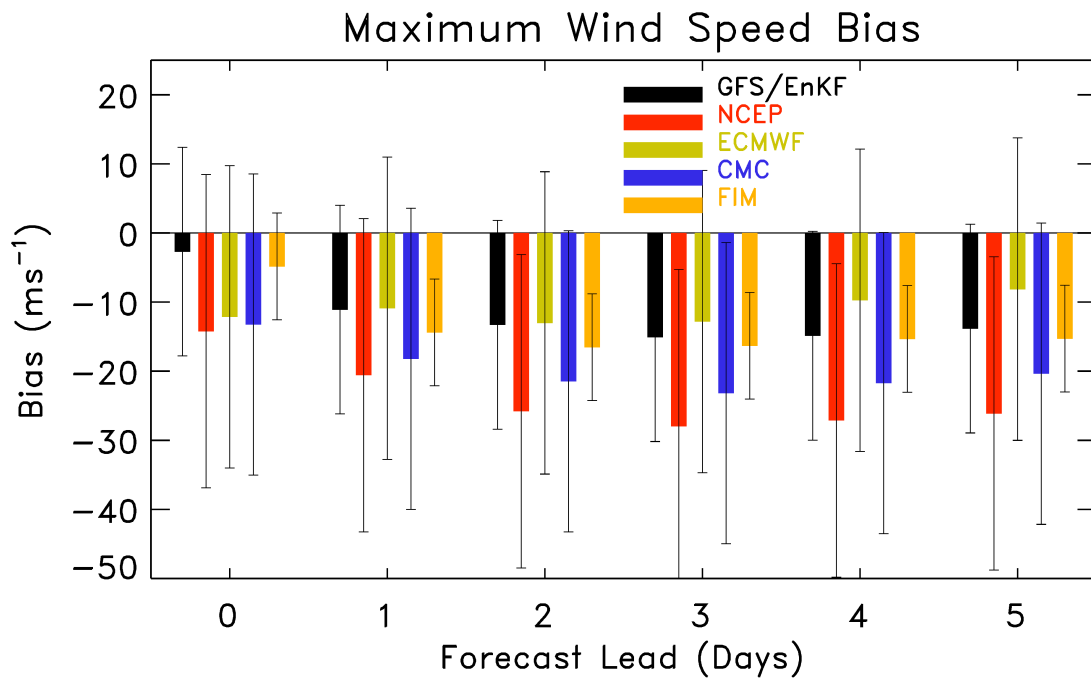


Fig. 13: Wind speed bias (average ensemble-mean forecast minus observed) for the forecast models used in this study (UKMO did not report wind speed). Error bars represent the 5th and 95th percentiles of a normal distribution fit to the available samples. Data here is non-homogeneous.

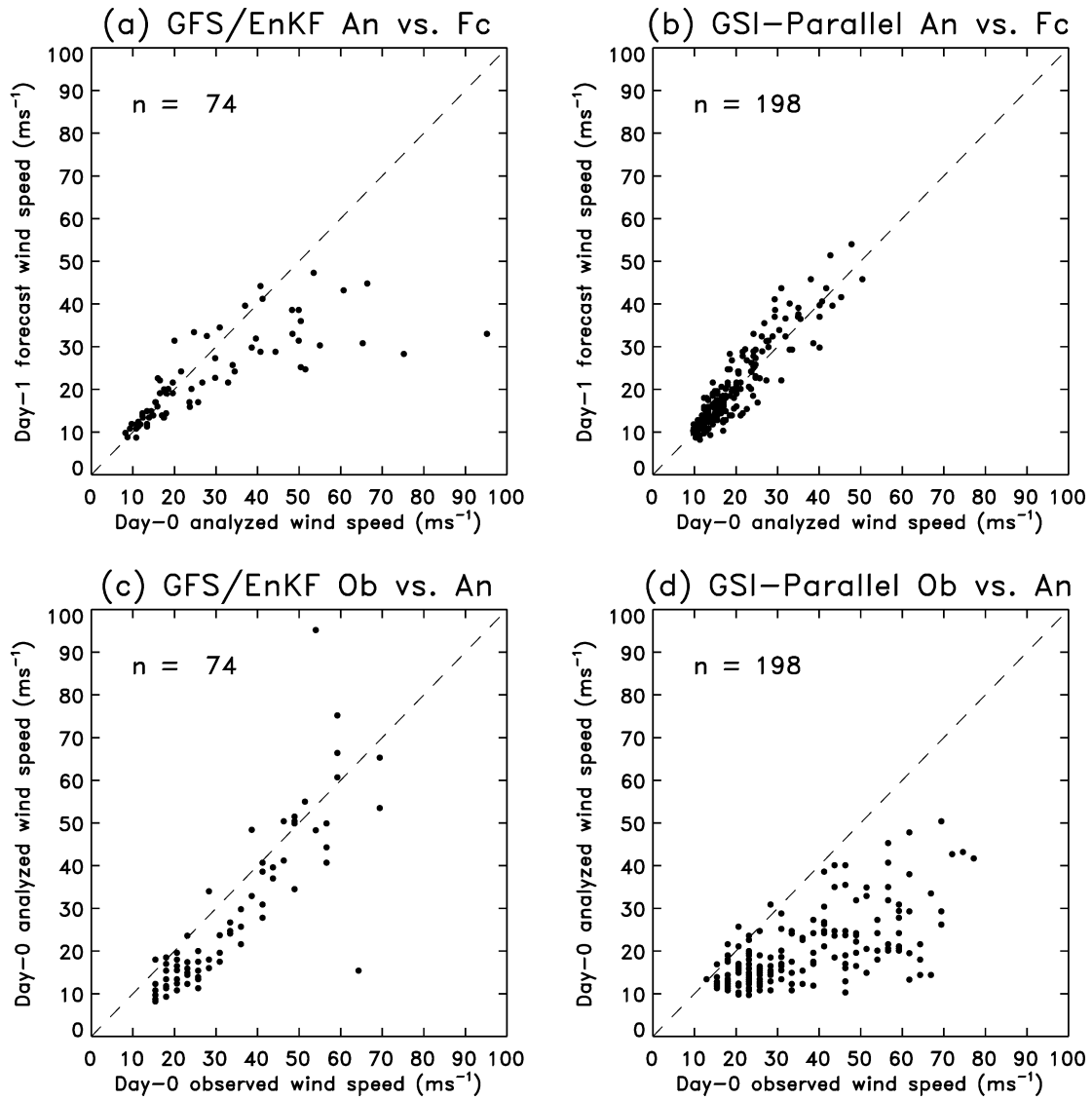
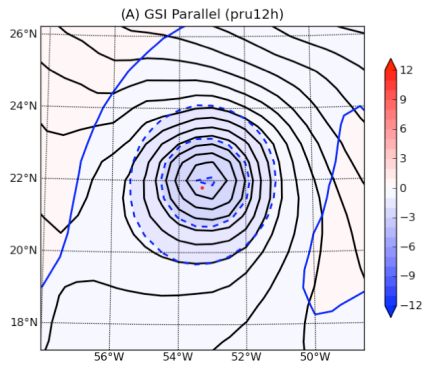


Fig. 14: Scatterplots of analyzed maximum wind speed (abscissa) vs. 1-day forecast wind speed (ordinate) for (a) GFS/EnKF deterministic forecast from ensemble-mean initial condition, and (b) GSI-parallel deterministic forecast. (c) Observed wind speed from TCVitals (abscissa) vs. analyzed wind speed (ordinate) for GFS/EnKF, and (d) GSI-Parallel.



F

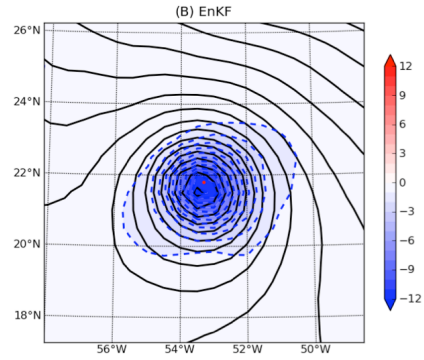


Fig. 15: Background forecast (black contours; interval 1 hPa) of sea-level pressure, and mean analysis increment (dashed lines and colors) for hurricane Ike at 0000 UTC 4 September 2008. (a) Parallel run of T382 GSI, and (b) T382 GFS.

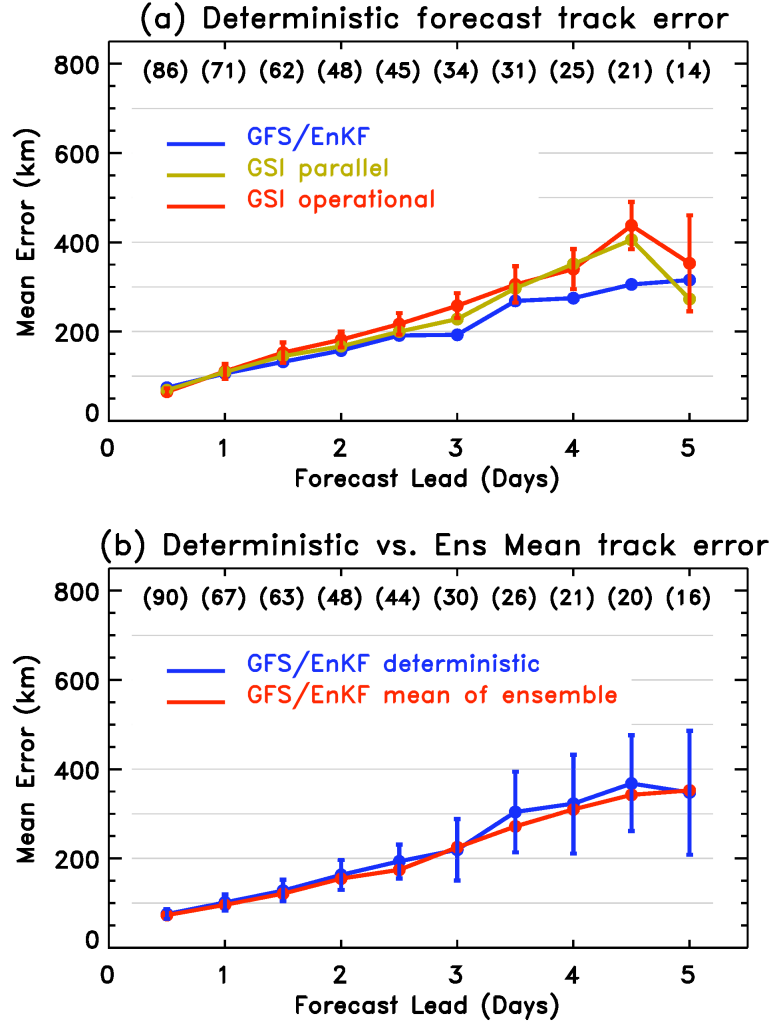


Fig. 16: (a) Homogeneous track error comparison for deterministic forecasts initialized from the GFS/EnKF ensemble-mean initial condition (blue), the GSI-parallel (yellow-green) and the GSI operational (red). Error bars represent the 95th percentiles of a paired block bootstrap distribution between GSI operational and GSI parallel. (b) Homogeneous track error comparison for deterministic forecasts initialized from the GFS/EnKF ensemble-mean initial condition (blue) vs. the mean of the ensemble-forecasts positions from the experimental GFS/EnKF (red).

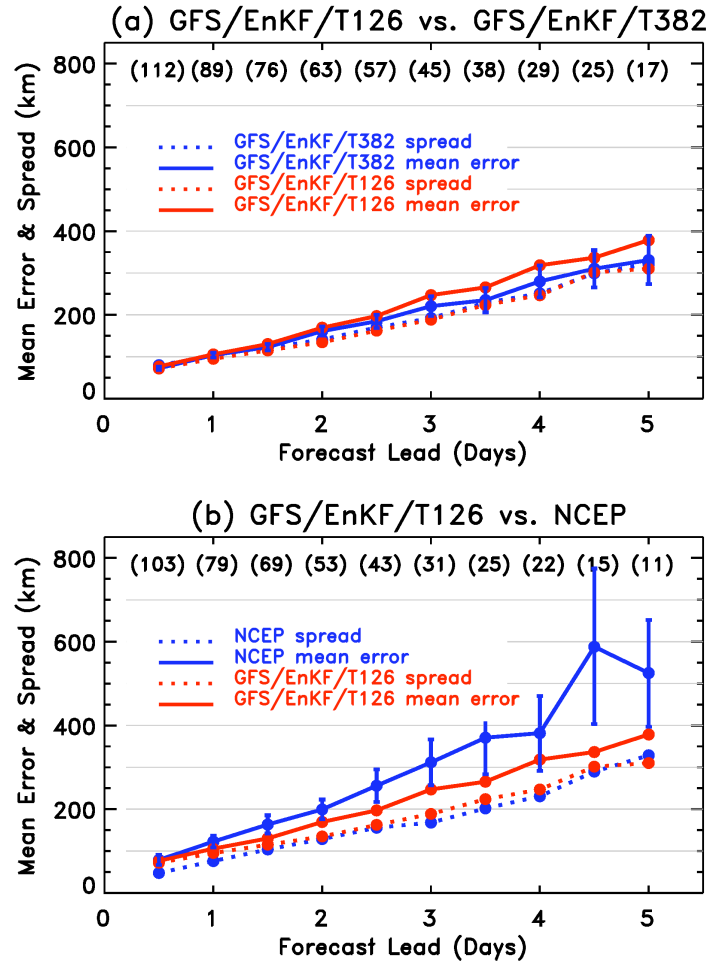


Fig. 17: Track error and spread, as in Fig. 5, but for (a) comparison of experimental GFS/EnKF with ensemble forecasts conducted at T382L64 resolution vs. experimental GFS/EnKF with ensemble forecasts conducted at the resolution of the 2009 operational system T126L28 (though initial conditions from T382L64 EnKF). (b) Comparison of operational T126L28 NCEP ensemble forecast system track errors to the experimental GFS/EnKF at the same resolution.

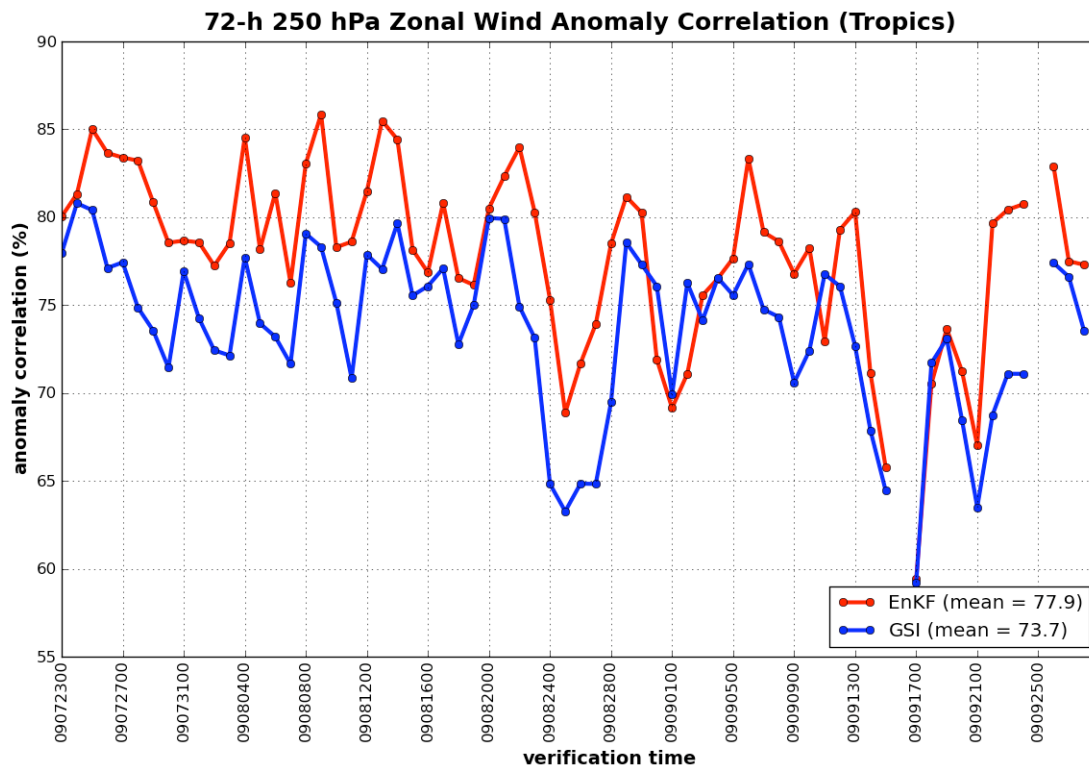


Fig. 18. Anomaly correlation of tropical (equatorward of 20 degrees) 250-hPa zonal wind from deterministic 72-h T382L64 GFS forecasts from the operational GSI analysis (blue) and the experimental GFS/EnKF analysis (red).